



Машинное обучение для решения исследовательских и инженерных задач в науках о Земле

Михаил Криницкий

к.т.н., Н.С.

Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и
мониторинга климатических изменений (ЛВОАМКИ)





Введение.

Михаил Криницкий

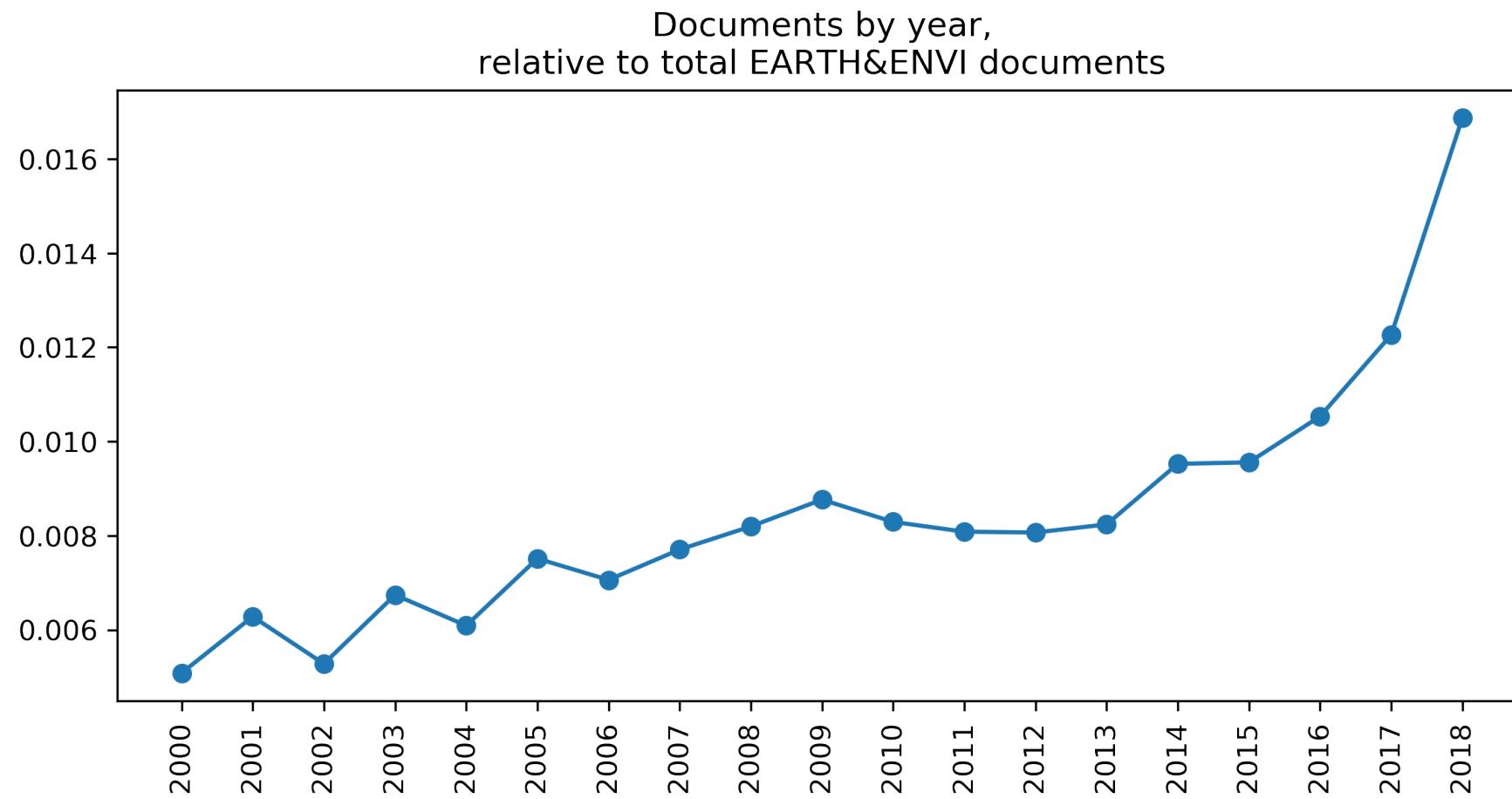
к.т.н., н.с.

Институт океанологии РАН им. П.П. Ширшова

Лаборатория взаимодействия океана и атмосферы и
мониторинга климатических изменений (ЛВОАМКИ)

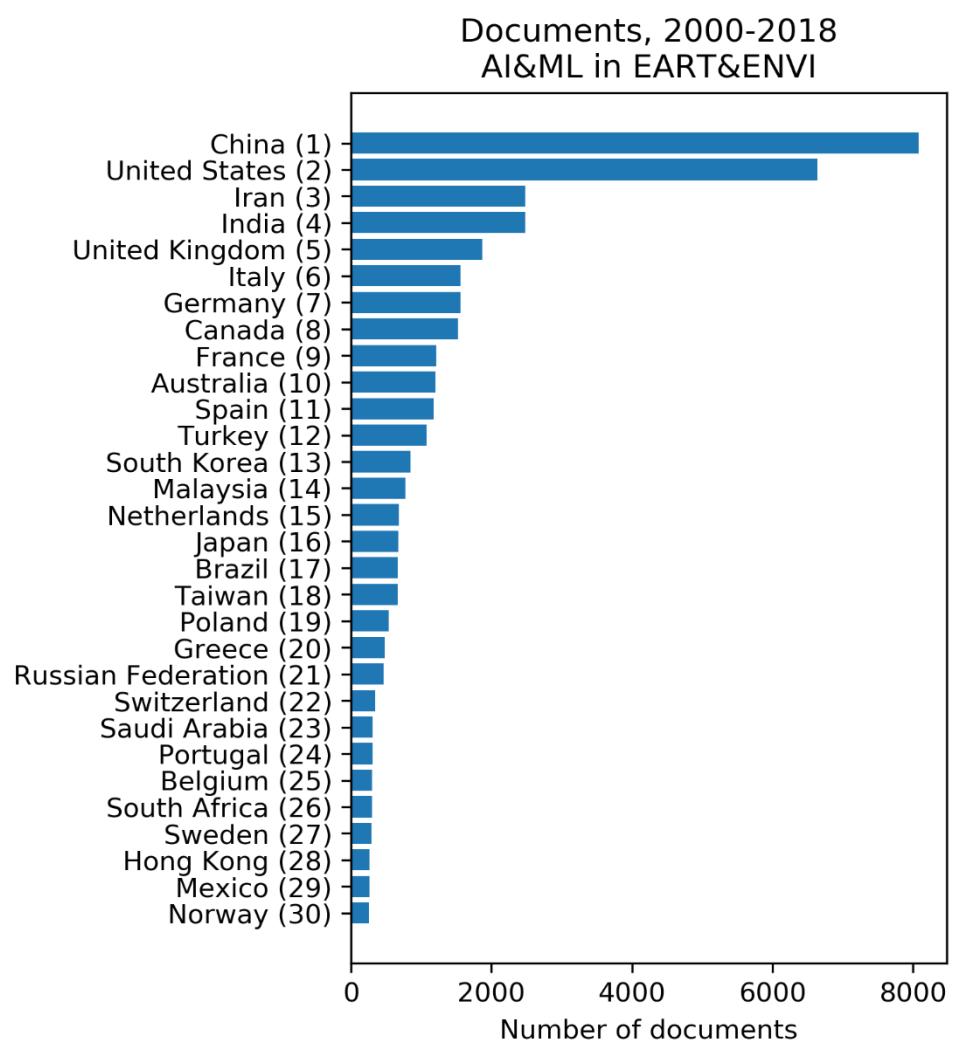
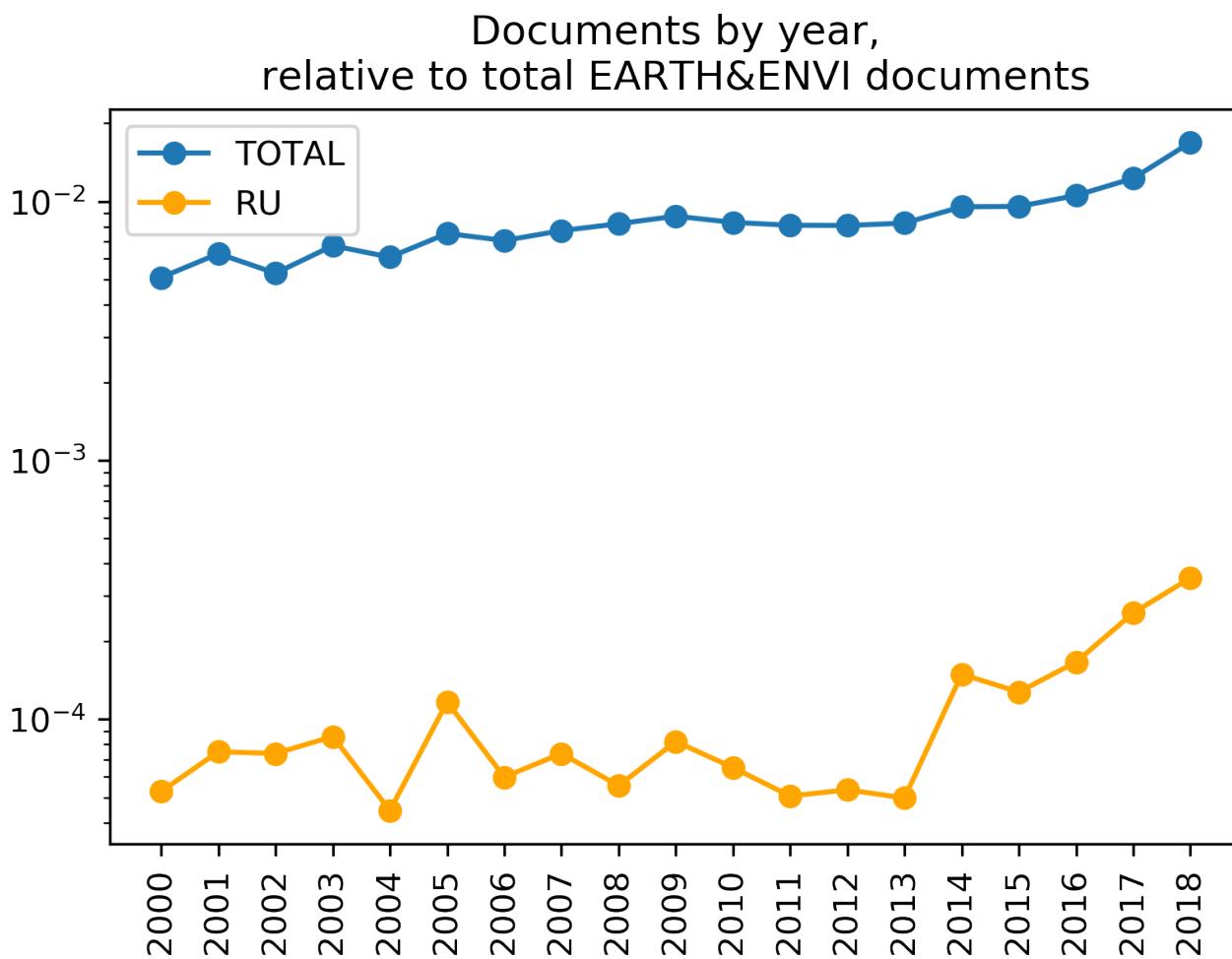
МО в науках о Земле

Статьи в рецензируемых журналах по тематике “Environmental sciences” и “Earth sciences” с применением методов машинного обучения. По данным Scopus.



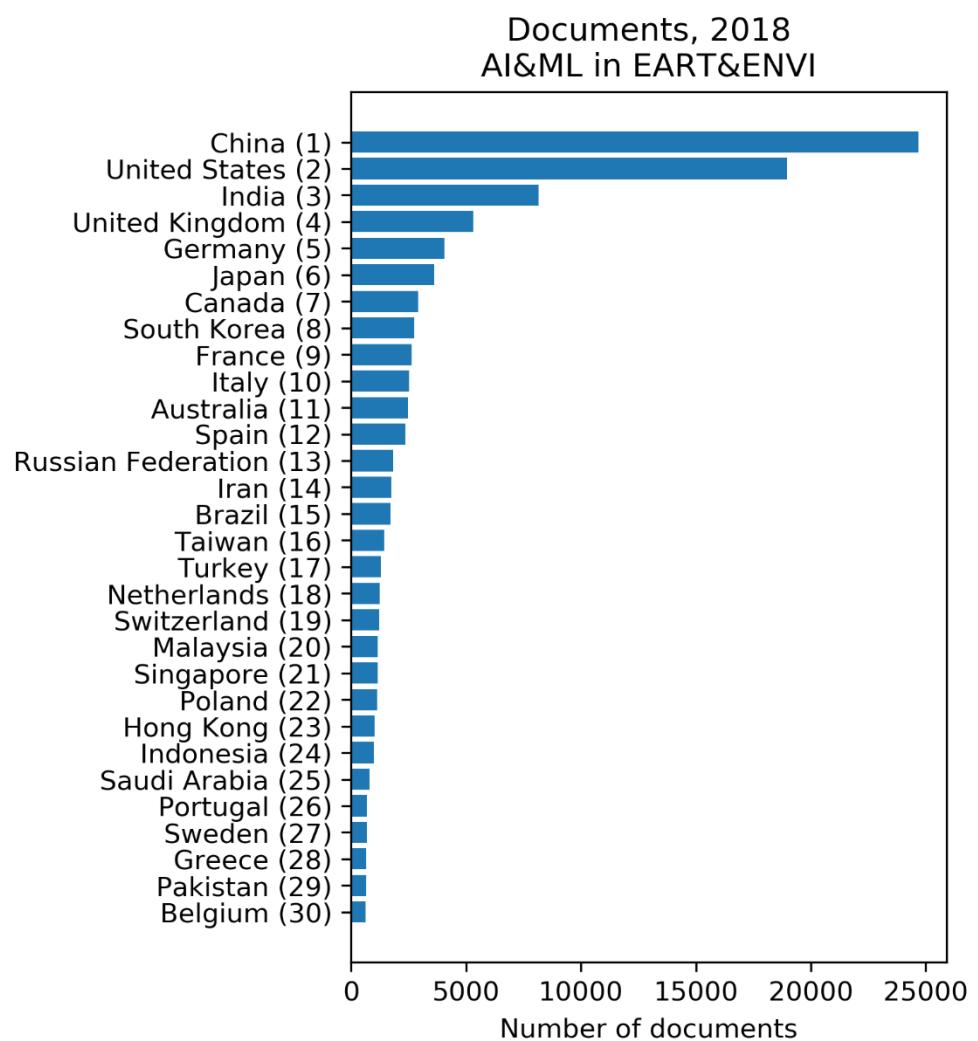
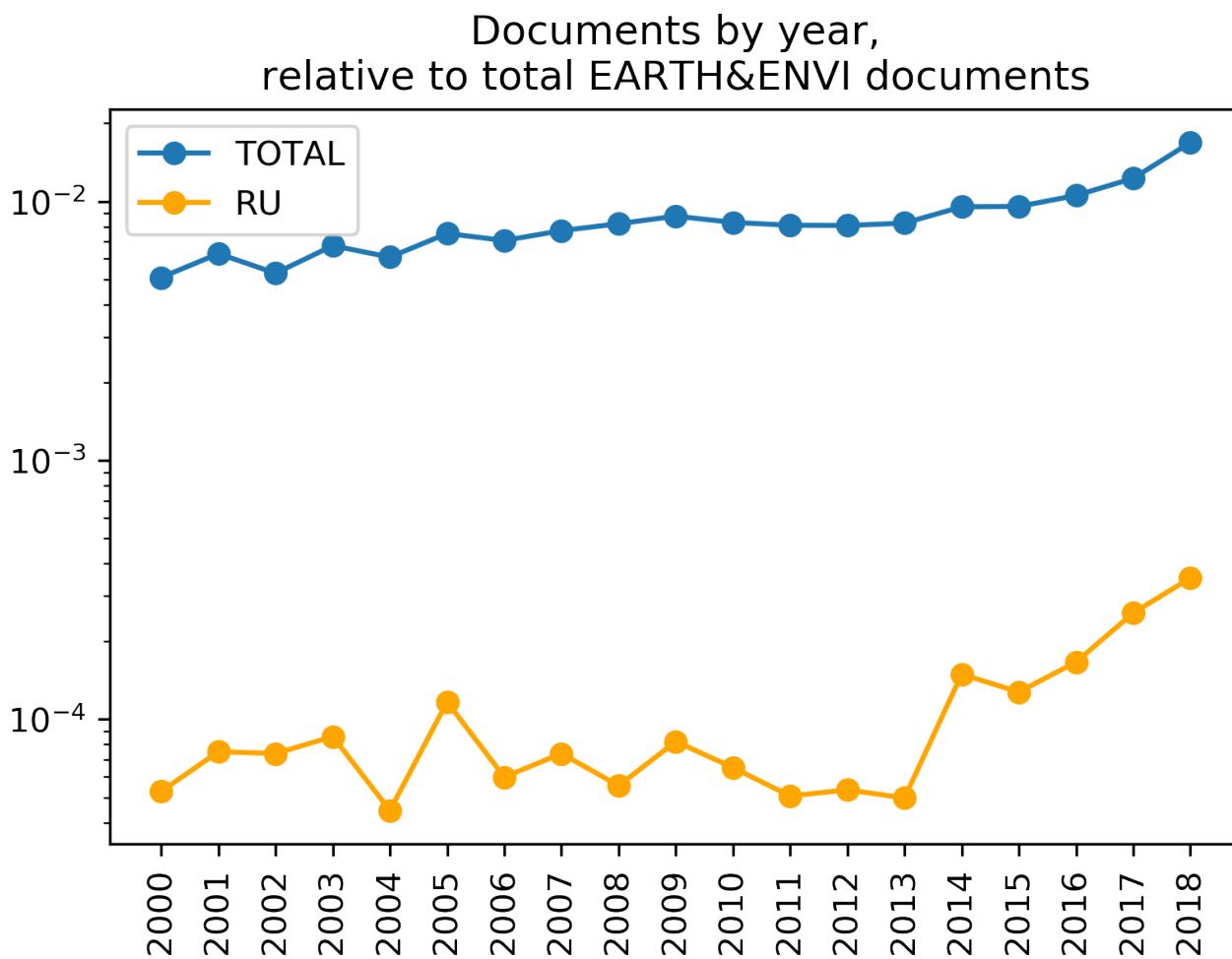
МО в науках о Земле

Статьи в рецензируемых журналах по тематике “Environmental sciences” и “Earth sciences” с применением методов машинного обучения. По данным Scopus.



МО в науках о Земле

Статьи в рецензируемых журналах по тематике “Environmental sciences” и “Earth sciences” с применением методов машинного обучения. По данным Scopus.



План лекции:

- Очень краткое введение в методы машинного обучения (ММО)
 - ММО: что это и зачем это нужно
 - Если я знаю, чего я хочу: «Обучение с учителем»
 - Если я еще не знаю, чего хочу: «Обучение без учителя»
- Примеры задач наук о Земле, решаемые с применением ММО:
 - Идентификация и построение траекторий полярных мезоциклонов
 - Кластеризация состояний стратосферного полярного вихря
 - Наукастинг атмосферных осадков...

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

ЧЕМ ЗАНИМАЮТСЯ ФИЗИКИ?

атмосферы

океана

взаимодействия океана и атмосферы

цунами

ядерщики, высоких энергий

био-

...

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

ЧЕМ ЗАНИМАЮТСЯ ФИЗИКИ?

ИЩУТ ИСТИНУ (об устройстве Вселенной) ?

ОТКРЫВАЮТ ЗАКОНЫ МИРОЗДАНИЯ ?

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

ЧЕМ ЗАНИМАЮТСЯ ФИЗИКИ?

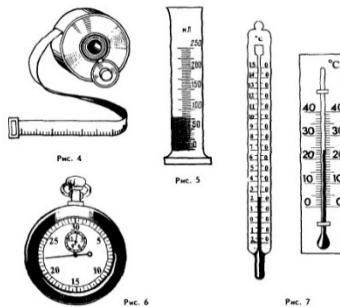
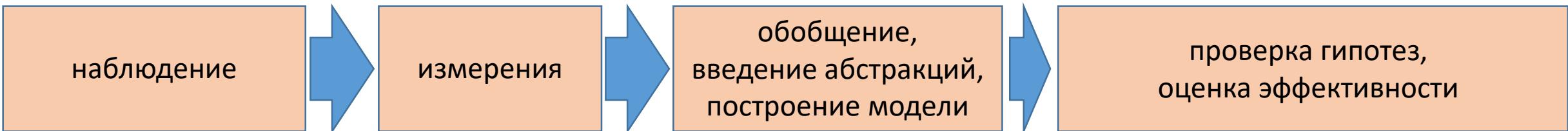
ПРЕДЛАГАЮТ МОДЕЛИ наблюдаемых процессов (**какие?..**)

ПРОВЕРЯЮТ - ОЦЕНИВАЮТ ЭФФЕКТИВНОСТЬ (**как?..**)

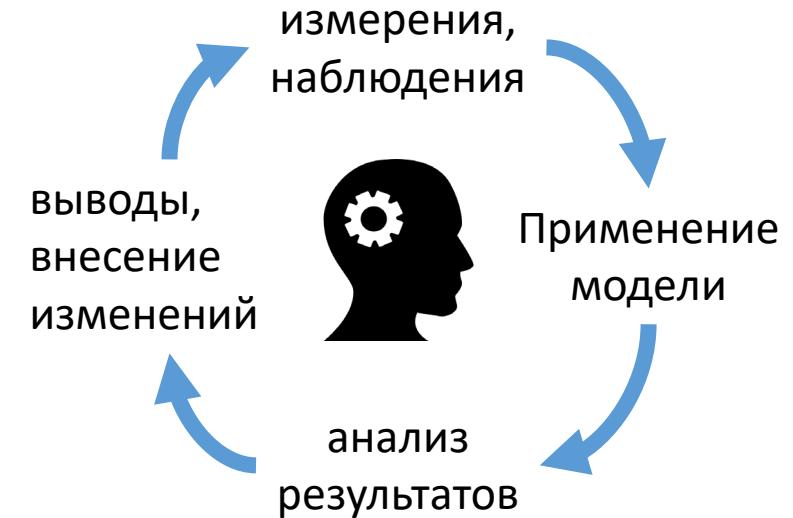
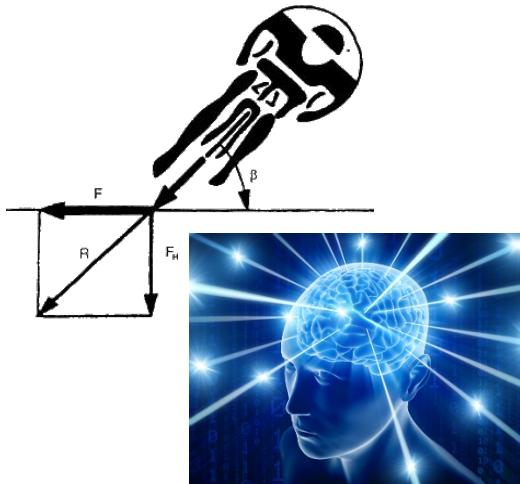
ПРИМЕНЯЮТ НА ПРАКТИКЕ

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

КАК проводятся физические исследования?



Настоящая наука начинается с тех
пор, как начинают измерять.
Точная наука немыслима без меры.
Д.И. Менделеев



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
но делать-то что-то надо

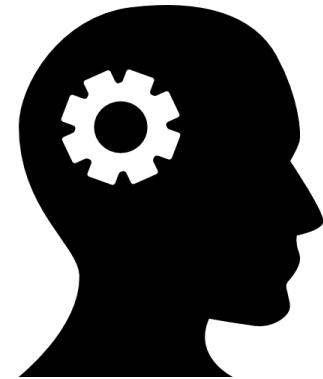
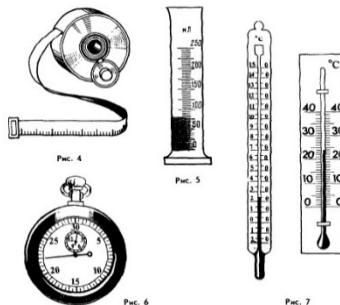
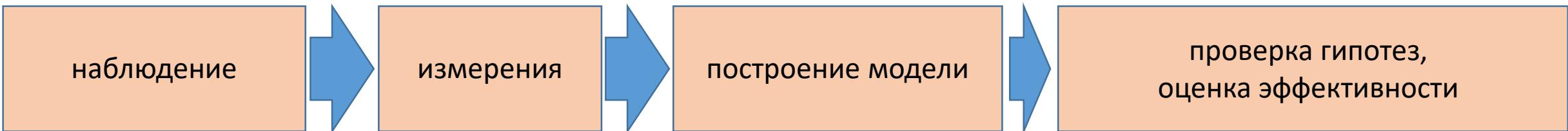
наблюдение



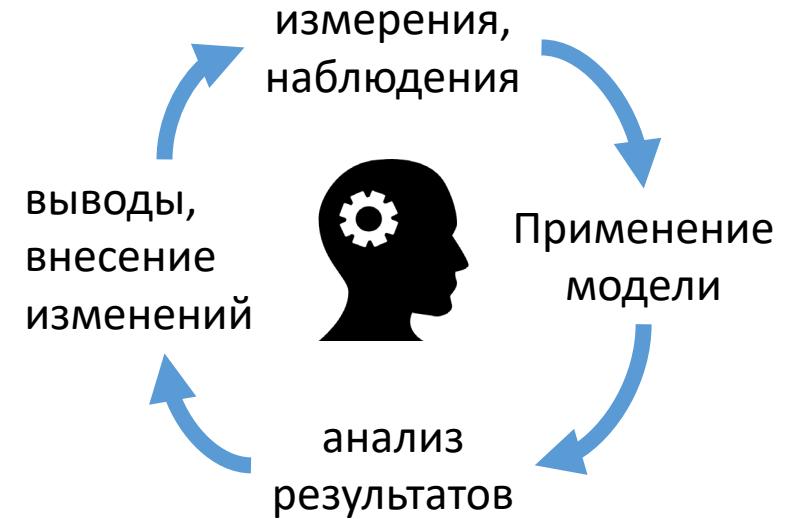
by Frits Ahlefeldt

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель



обобщение ?
введение абстракций ?



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы) $\{y_i\}$
-
-
-

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы) $\{y_i\}$
- Данные измерений $\{x_i\}$ (желательно
релевантных задаче)

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы) $\{y_i\}$
- Данные измерений $\{x_i\}$
- $\mathcal{F}: \mathbb{X} \rightarrow \mathbb{Y}$

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель

- Для чего? Какова цель?
- Что у нас для этого есть?
- Какого рода модель?
- Какая должна быть модель?
- Оценить неизвестную(ые) величину(ы) $\{y_i\}$
- Данные измерений $\{x_i\}$
- $\mathcal{F}: \mathbb{X} \rightarrow \mathbb{Y}$
- Обобщающая. Достоверная (в каком смысле?)

Применимая.

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель

КАК?

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Когда (человеку) непонятно, что происходит
все равно строим модель

КАК?

Методы машинного обучения

Искусственный интеллект

Теория Вапника-Червоненкиса

Статистическая теория восстановления
зависимостей по эмпирическим данным

Машинный интеллект

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач МО

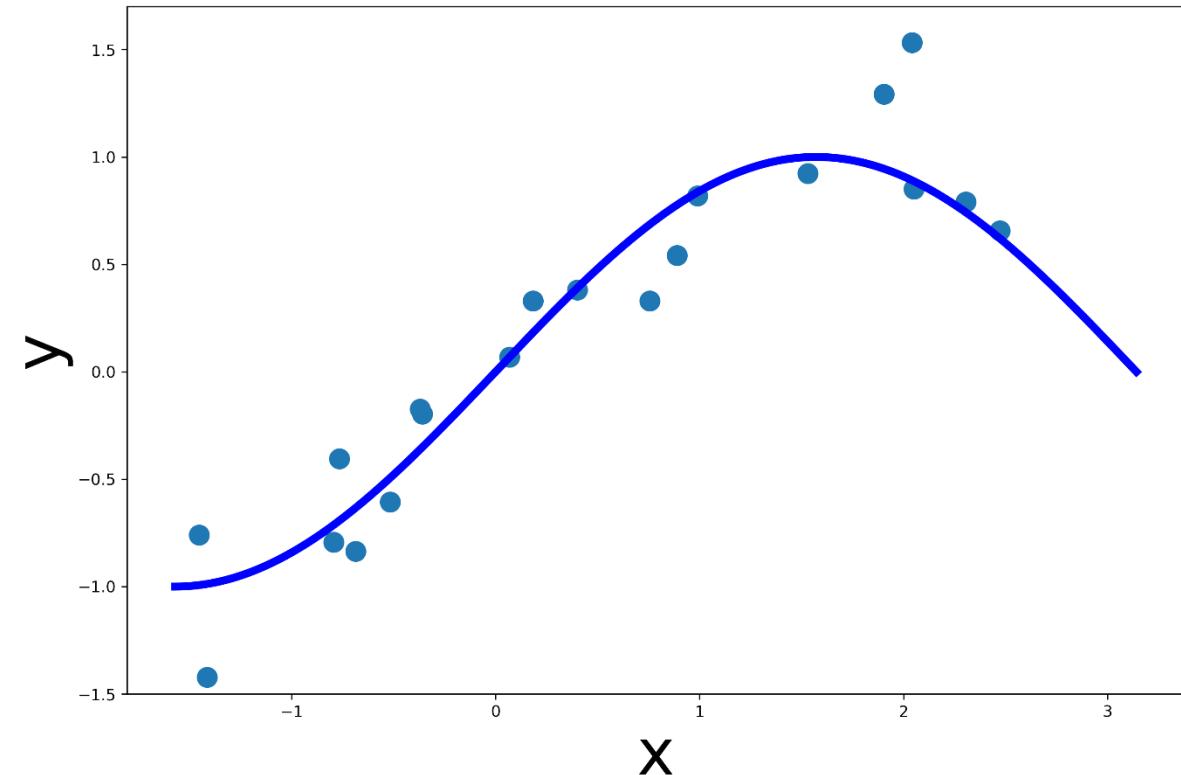
ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

строим модель для решения задачи

типы задач:

- «Обучение с учителем»
 - восстановление регрессии

что я хочу? – значение y



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

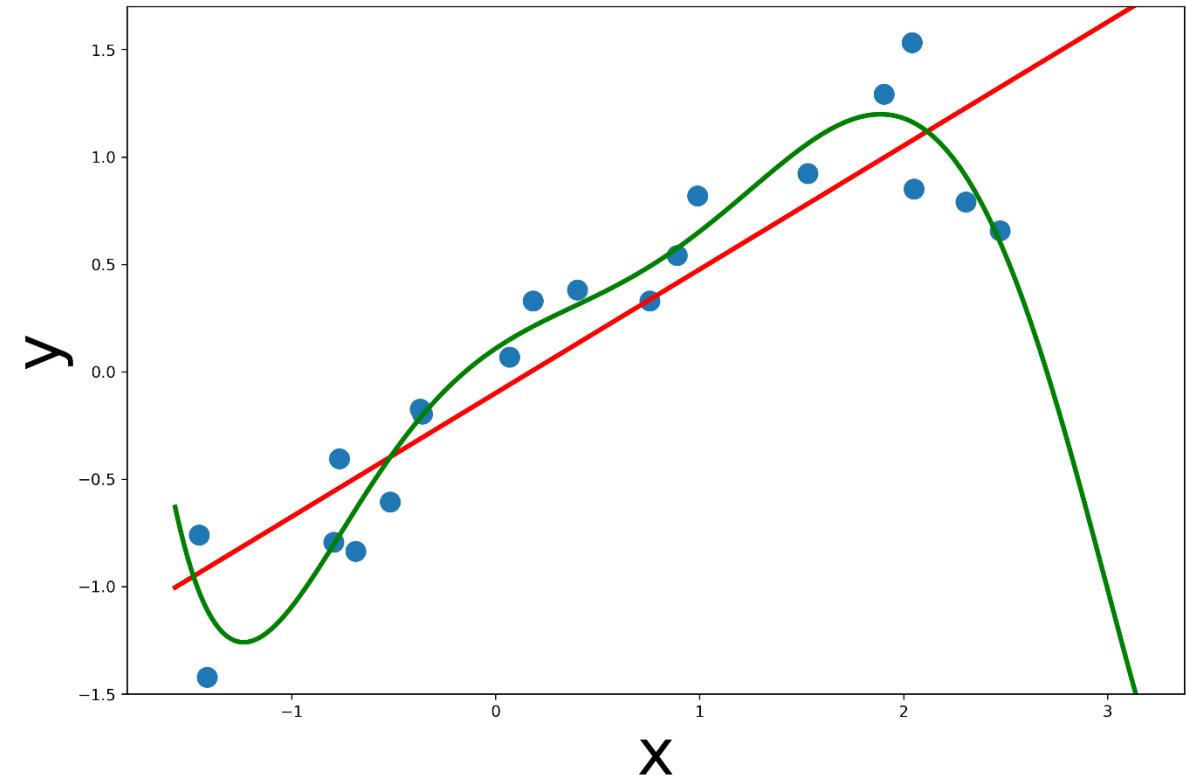
строим модель для решения задачи

типы задач:

- «Обучение с учителем»
 - восстановление регрессии

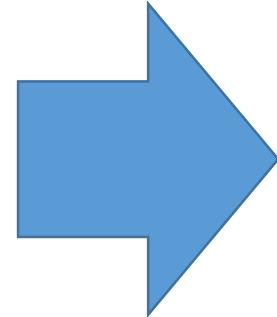
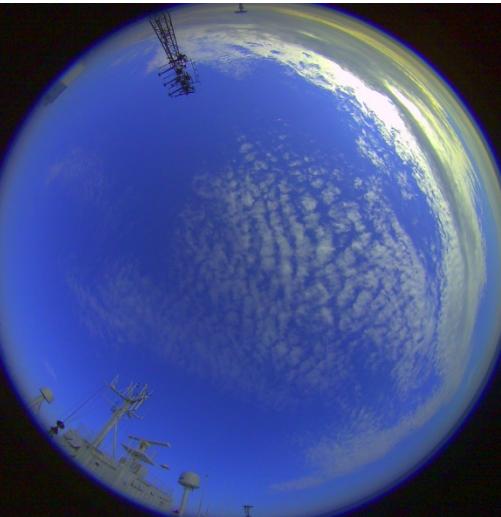
$$\hat{y} = ax + b$$

$$\hat{y} = p^{(6)}(x)$$



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

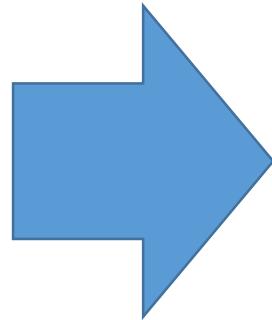
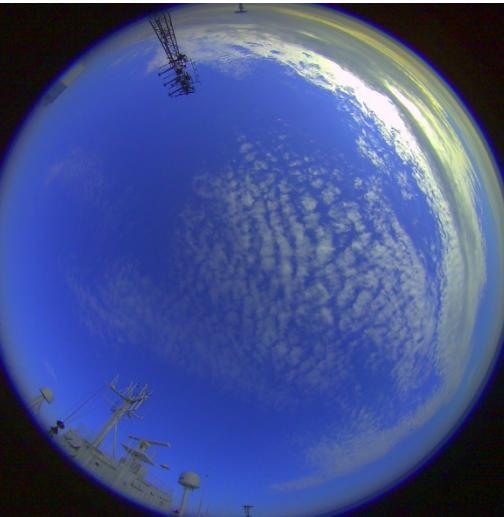
пример



y - Балл общей облачности
(0-8 баллов)

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

пример

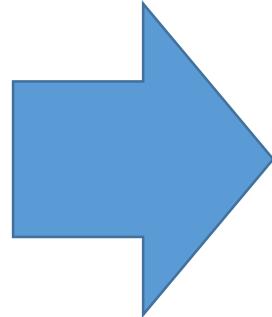
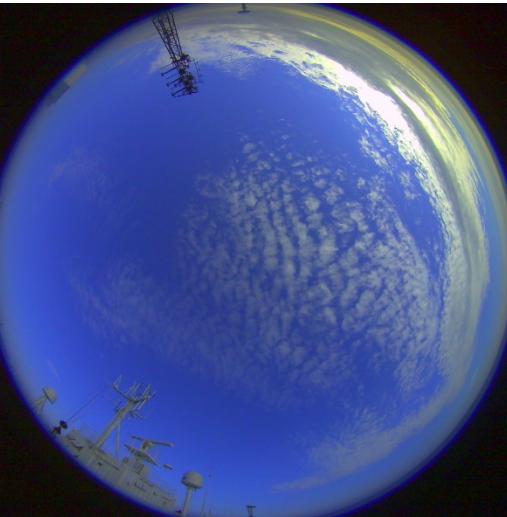


y - Балл общей облачности
(0-8 баллов)

x_i (объект, событие) – облачная ситуация, наблюдалася в момент снимка

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

пример



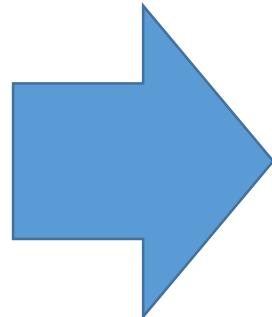
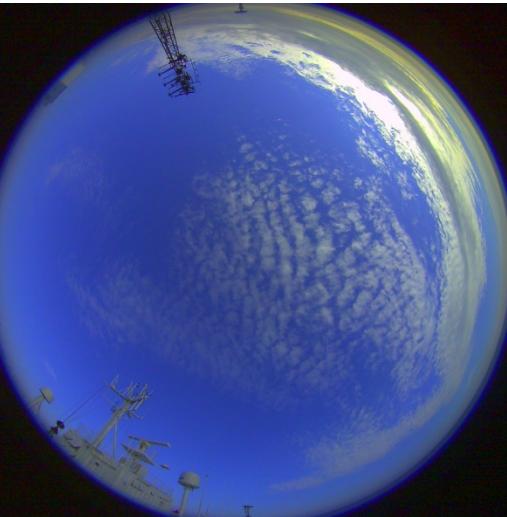
y - Балл общей облачности
(0-8 баллов)

x_i (объект, событие) – облачная ситуация, наблюдалася в момент снимка

\vec{x}_i (признаковое описание события) – яркость, осредненная по снимку

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

пример



y - Балл общей облачности
(0-8 баллов)

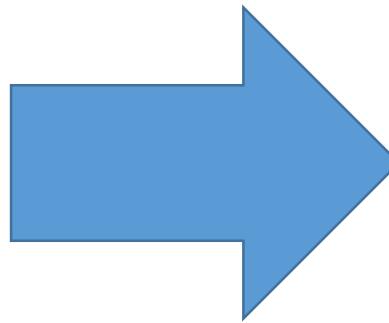
x_i (объект, событие) – облачная ситуация, наблюдаемая в момент снимка

\vec{x}_i (признаковое описание события) – медиана значений яркости на снимке

y_i – известный балл общей облачности для объектов обучающей выборки

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

пример



y - Балл общей облачности
(0-8 баллов)

x_i (объект, событие) – облачная ситуация, наблюдалася в момент снимка

\vec{x}_i (признаковое описание события) – медиана значений яркости на снимке

y_i – известный балл общей облачности для объектов обучающей выборки

Модель в задаче восстановления регрессии: $\hat{y}_i = f(\vec{p}, \vec{x}_i)$

Функция потерь:

$$\mathcal{L}(\vec{p}, \{\vec{x}_i\}) = \frac{1}{n} \sum_{i=1}^n (f(\vec{p}, \vec{x}_i) - y_i)^2$$

ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

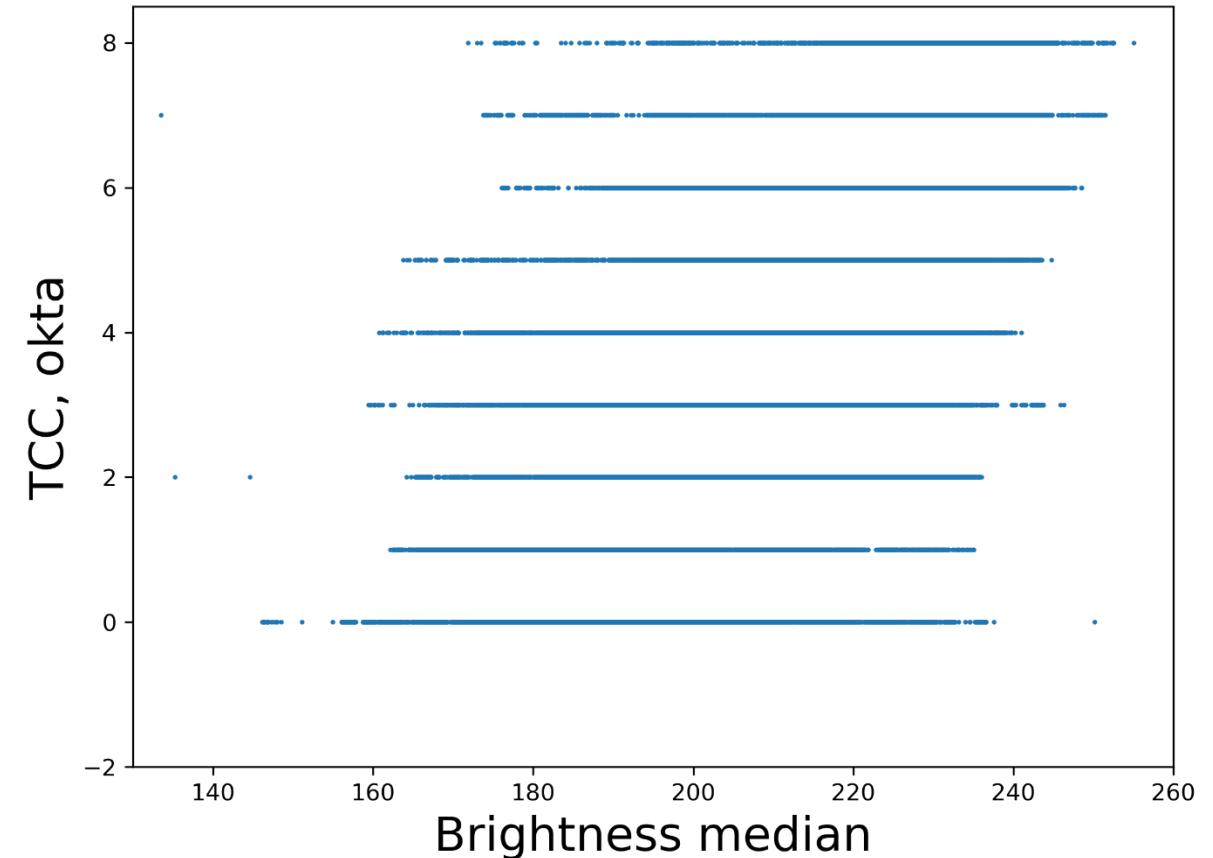
возьмем очень слабую модель

x_i - медиана значений яркости на всем снимке

$$\hat{y}_i = kx_i + b$$

Функция потерь:

$$\mathcal{L}(k, b, \{x_i\}) = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$



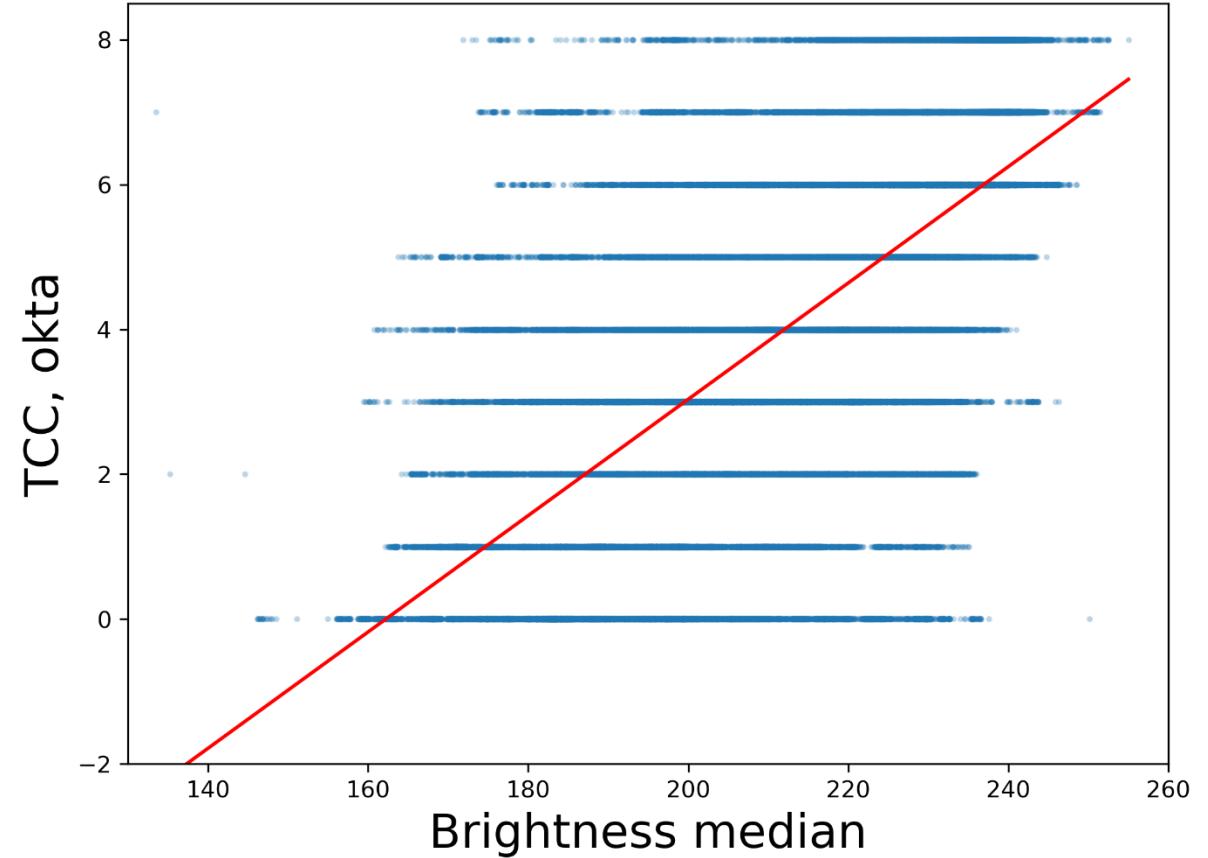
ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

x_i - медиана значений яркости по всему снимку

$$\hat{y}_i = kx_i + b$$

Функция потерь:

$$\mathcal{L}(k, b, \{x_i\}) = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

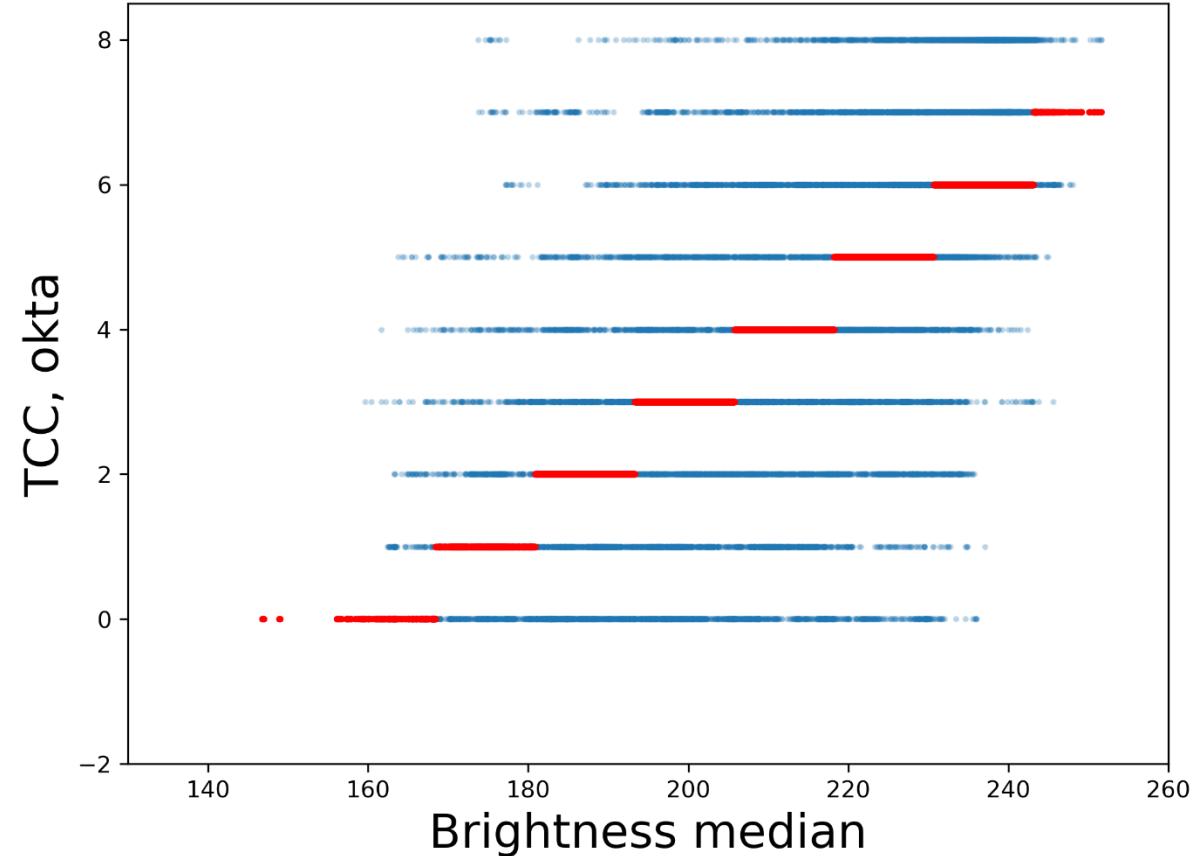
x_i - медиана значений яркости по всему снимку

$$\hat{y}_i = kx_i + b$$

Функция потерь:

$$\mathcal{L}(k, b, \{x_i\}) = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

Хорошая ли это модель?



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

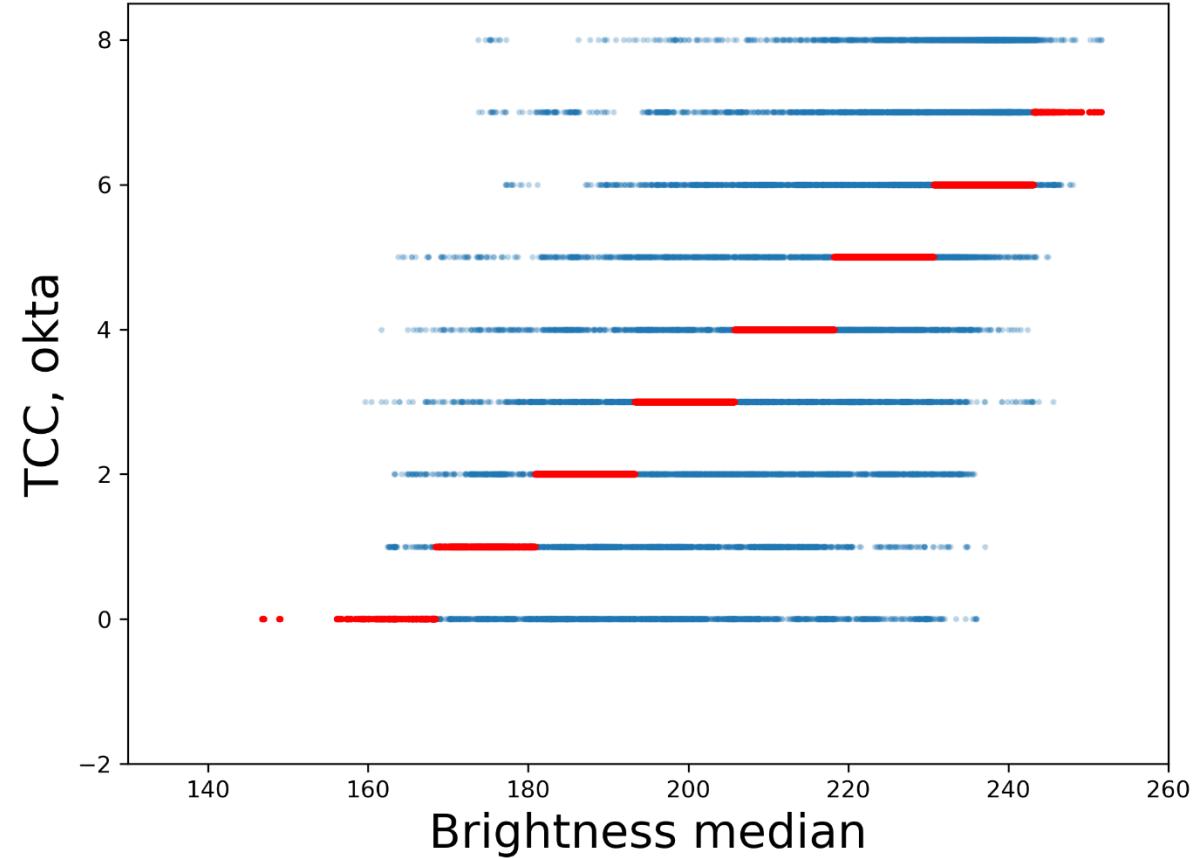
x_i - медиана значений яркости по всему снимку

$$\hat{y}_i = kx_i + b$$

Функция потерь:

$$\mathcal{L}(k, b, \{x_i\}) = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

Хорошая ли это модель?
зависит от того, как оценивать!



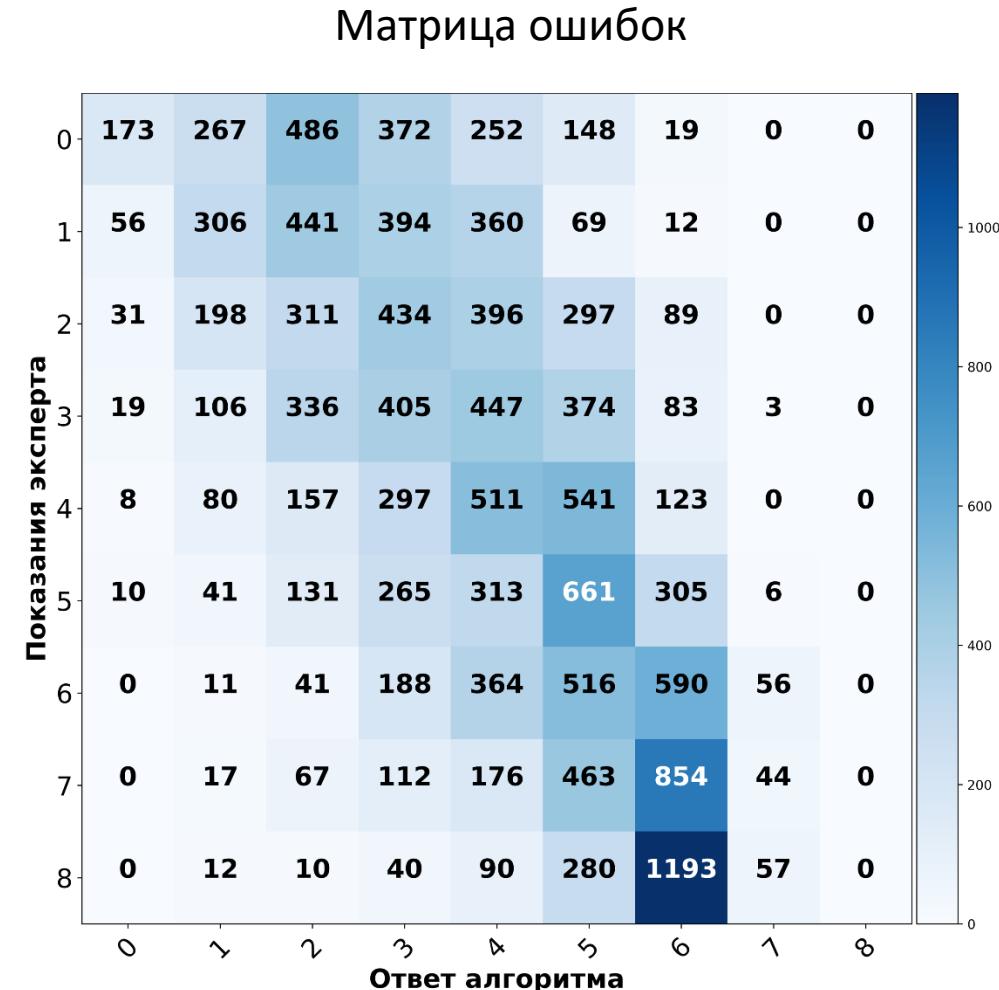
ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

метрики качества (в задаче оценки балла общей облачности)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} = 1.98 \text{ окта}$$

Accuracy, точность – доля верно оцененных значений

Acc = 19%

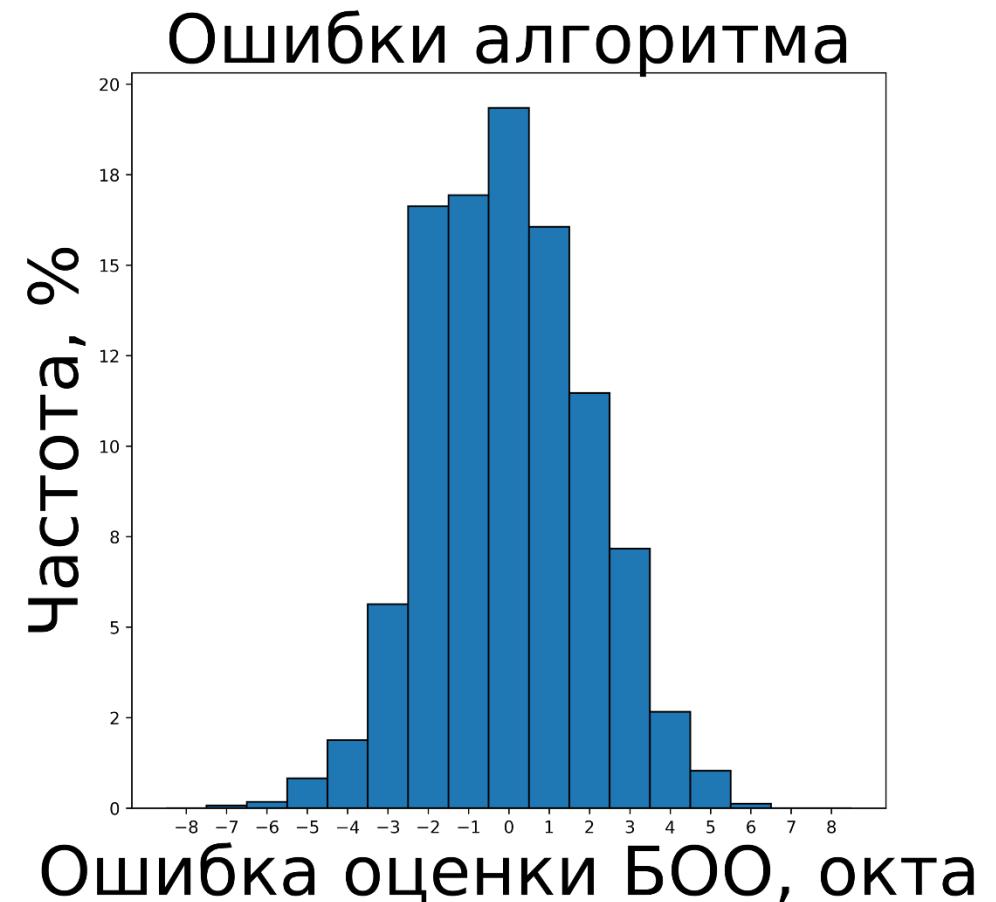


ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

метрики качества (в задаче оценки балла общей облачности)

Доля значений, оцененных с ошибкой не более 1 окта

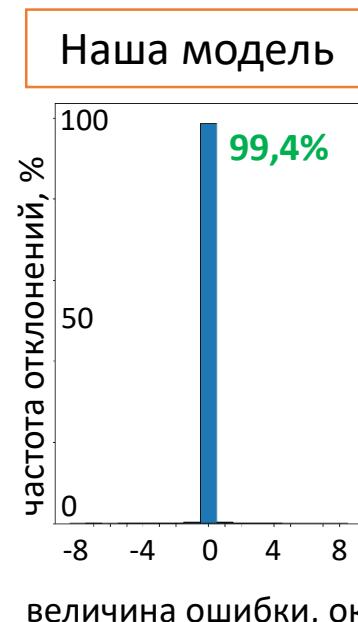
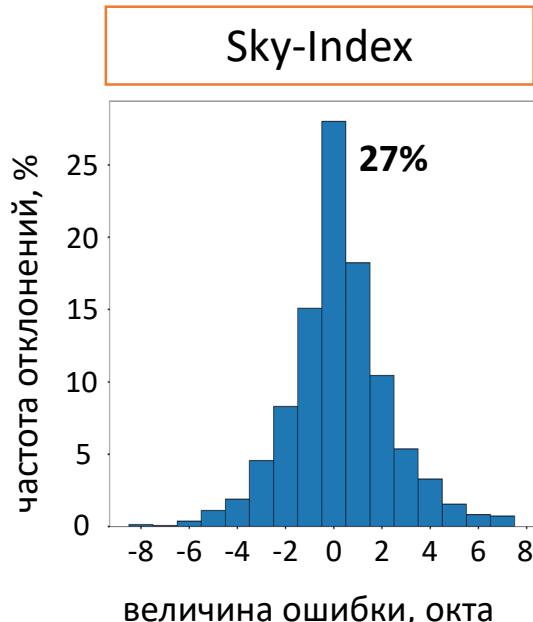
$$C_1 = 52 \%$$



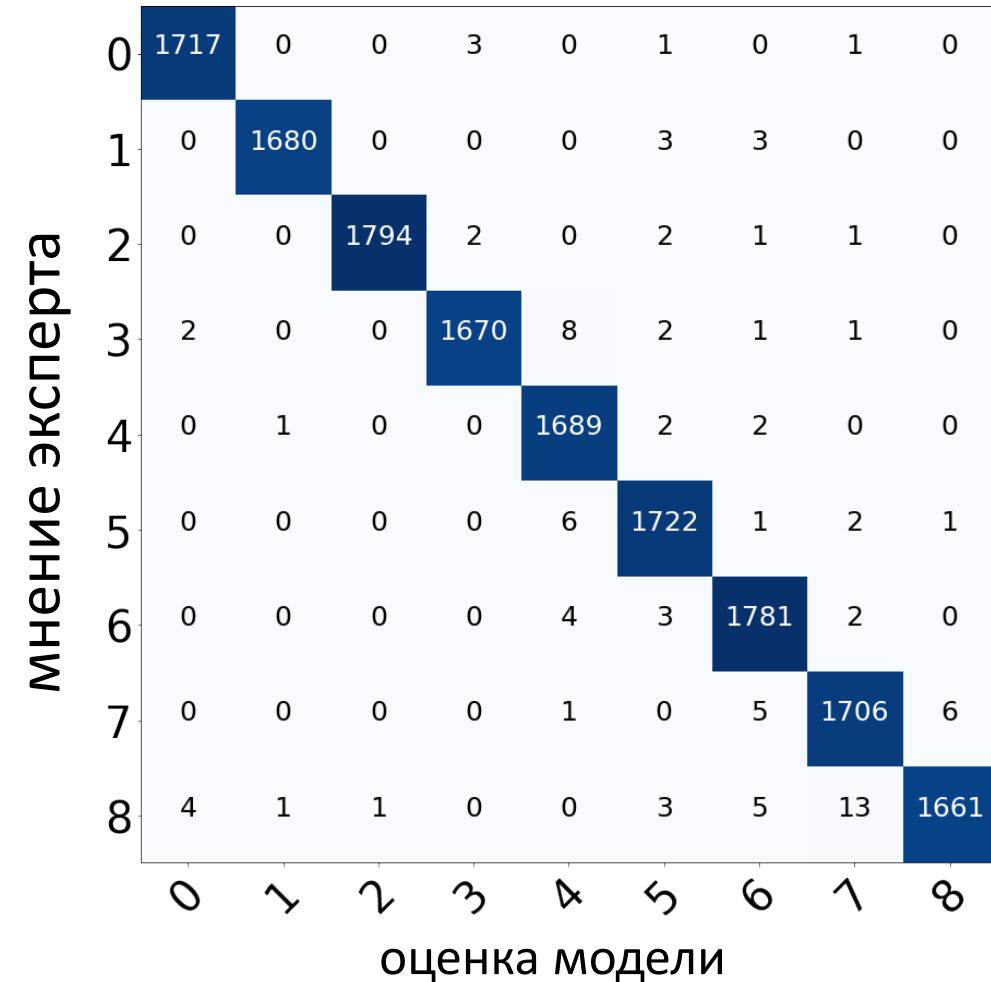
ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

	Sky-Index	Наша модель
<i>RMSE</i>	2.06	0.23
<i>Acc</i>	27%	99,4%
<i>C₁</i>	62%	99,7%

Ошибки алгоритмов

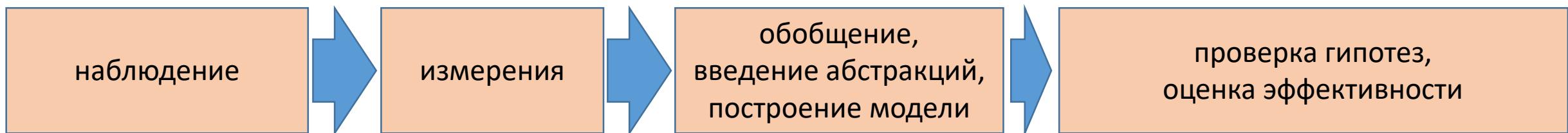


Матрица ошибок оценки ОБО

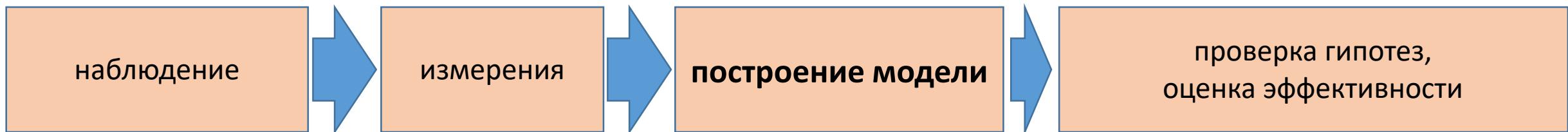


ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

классический подход



подход, основанный на данных



ОЧЕНЬ КРАТКОЕ ВВЕДЕНИЕ В МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

классический подход

- Трактуемость (интерпретируемость)
- В сложных случаях – долгая реализация
- Меняющиеся условия и неучтенные воздействия снижают точность
- Подход основан на понимании описываемого процесса. Модель процесса от начала до конца формулируется человеком.
поэтому нередко модель **недостаточно сложна** для описания сложного процесса с **желаемой точностью**.

методы машинного обучения

- Трактуемость (интерпретируемость) – **редко**
ММО не всегда подходят для установления причинно-следственных зависимостей
- Скорость реализации
- Модели можно «дообучать» в меняющихся условиях
- Подход основан на данных. Человек может участвовать (помогать) на этапе выбора типа модели общего вида.
модели МО обладают широким спектром сложности, что **позволяет аппроксимировать** очень сложные процессы с **хорошей точностью**.

ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Идентификация и построение траекторий полярных мезоциклонов (ПМЦ)

Михаил Криницкий¹, Полина Вереземская^{1,2}, Кирилл Гращенков^{1,3}, Иван Охлопков^{1,4}

¹ Институт океанологии им. П.П. Ширшова РАН, ЛВОАМКИ

² Научно-исследовательский вычислительный центр, МГУ им. М.В. Ломоносова

³ Московский физико-технический институт (государственный университет), ФАКИ.

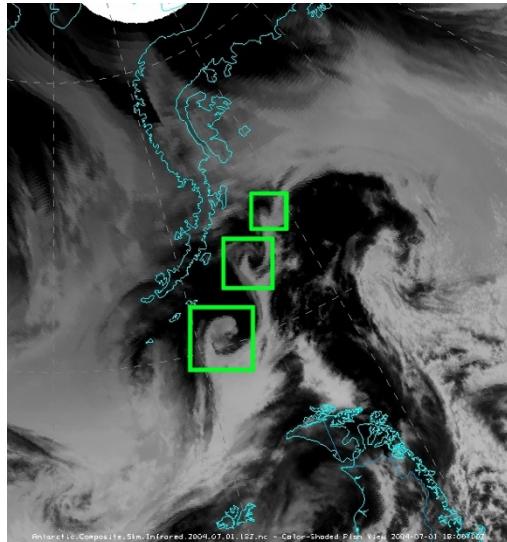
⁴ Московский государственный университет им. М.В. Ломоносова, Географический ф-т.

ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Идентификация и построение траекторий ПМЦ

Что такое ПМЦ?

Короткоживущие	up to 36 hours
Быстро развивающиеся	3-6 ч.
Небольшие	200-1000 km
Интенсивные	скорость ветра >15 m/s
В-основном зимнее и весеннеевремя	



Зачем?

(искать ПМЦ и строить траектории)

- Влияют на формирование глубокой конвекции в океане
- Несут угрозу судам и морским инфраструктурным объектам
- Слабо предсказуемы
- Слабо документированы, особенно в Южном Полушарии

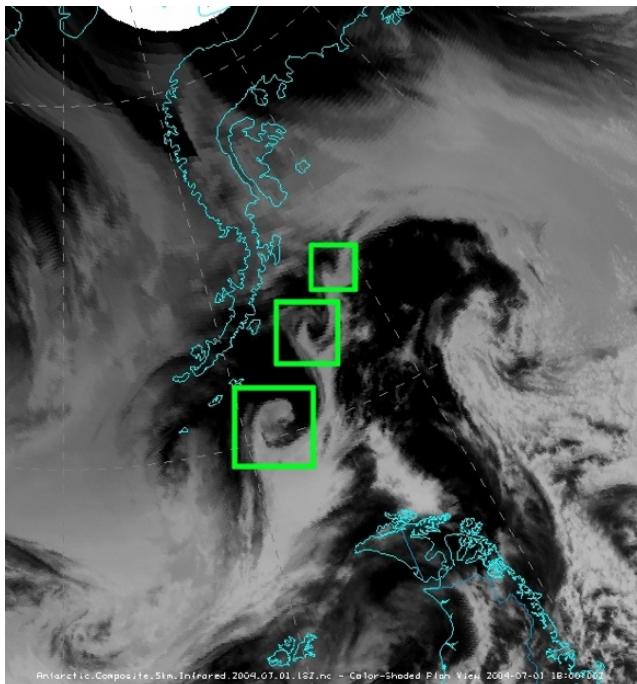
ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Идентификация и построение траекторий ПМЦ

Как это делают сейчас?

Исходные данные:

- Численное моделирование атмосферы (атмосферные реанализы)
- Спутниковые снимки



Способы трекинга:

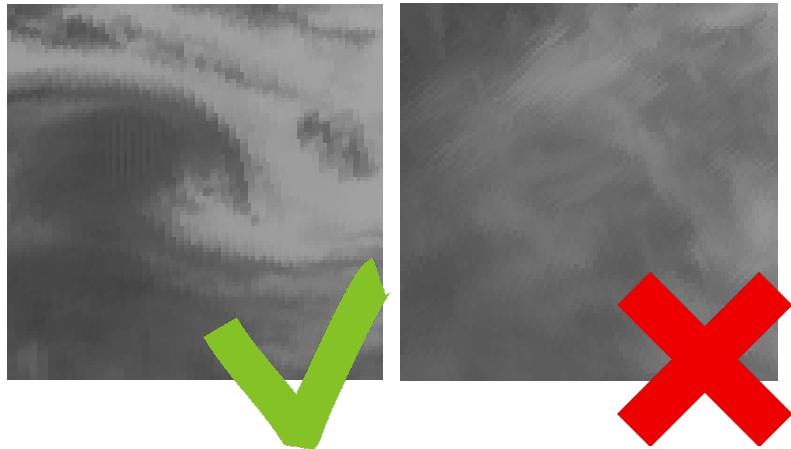
- Автоматический (фильтрация, применение пороговых схем – на данных численного моделирования)
- Визуальный (по спутниковым данным)

ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Идентификация и построение траекторий ПМЦ

Как это делаем мы?

Задача №1 - классификация

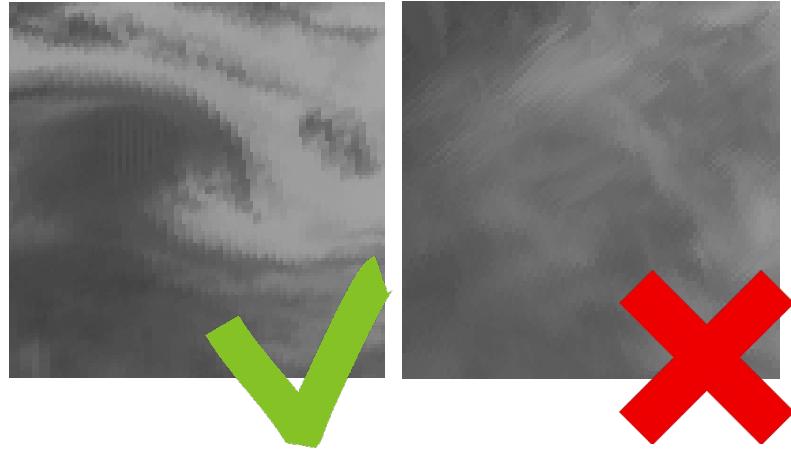


ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

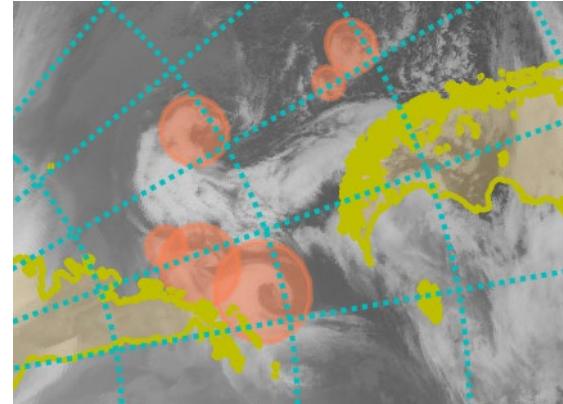
Идентификация и построение траекторий ПМЦ

Как это делаем мы?

Задача №1 - классификация



Задача №2 - сегментация

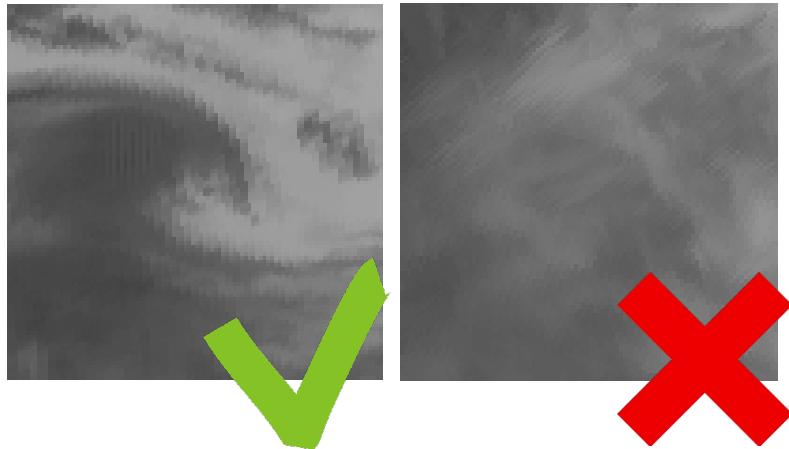


ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

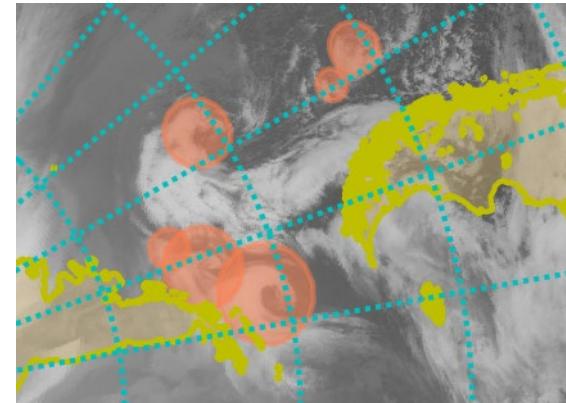
Идентификация и построение траекторий ПМЦ

Как это делаем мы?

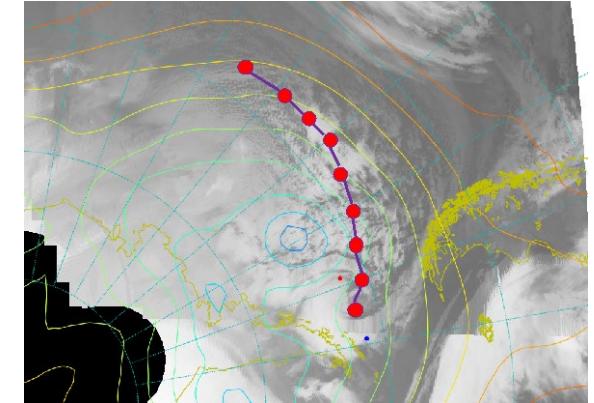
Задача №1 - классификация



Задача №2 - сегментация



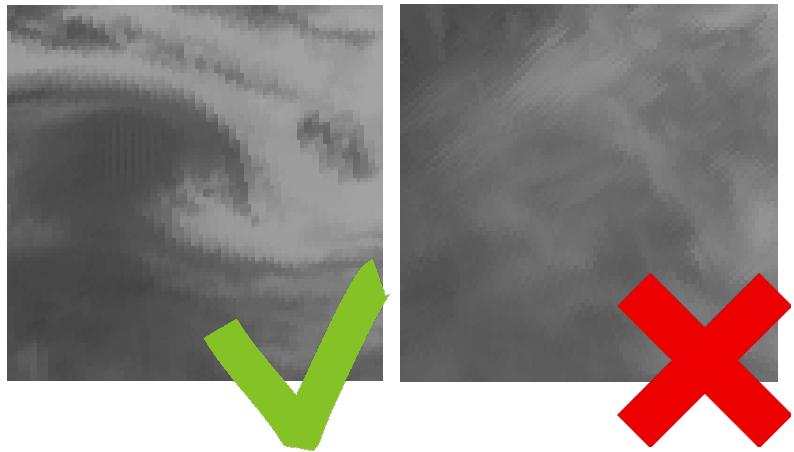
Задача №3 - трекинг



ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Идентификация и построение траекторий ПМЦ

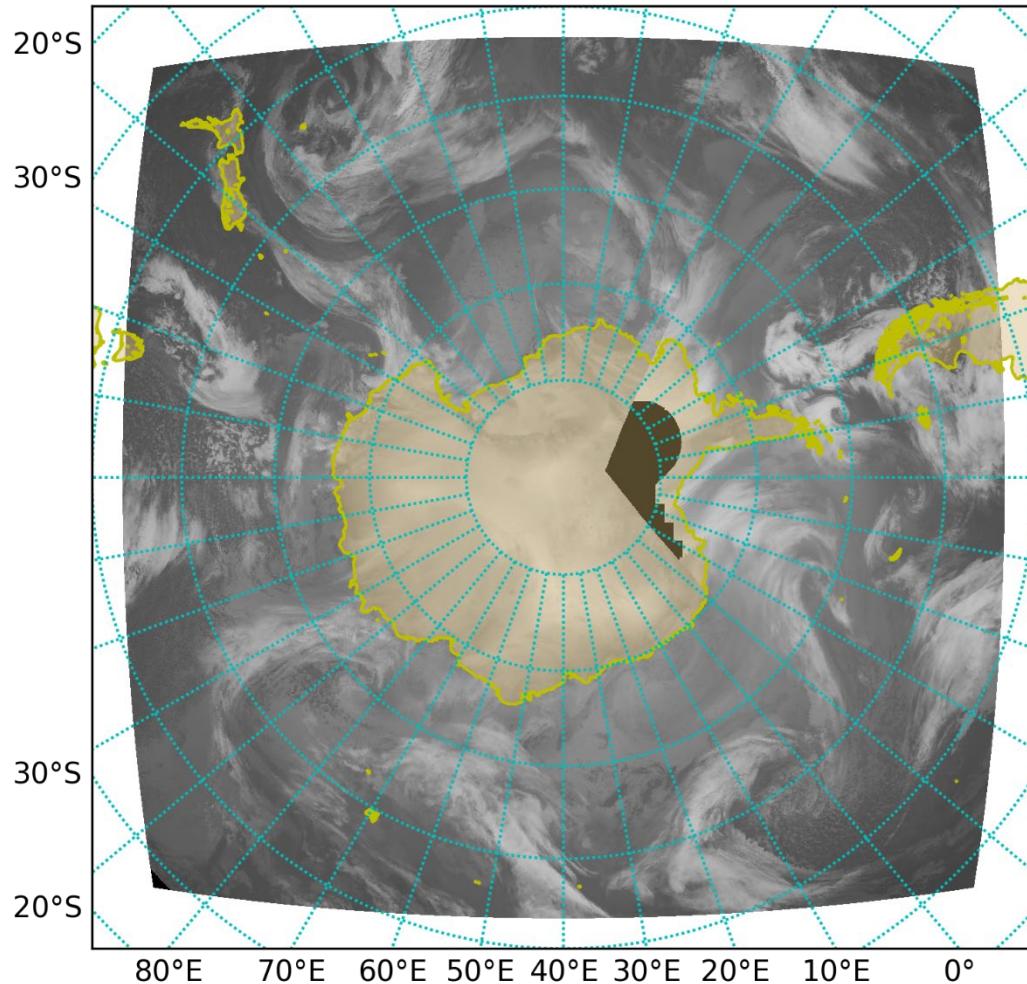
Задача №1 - классификация



ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Идентификация и построение траекторий ПМЦ

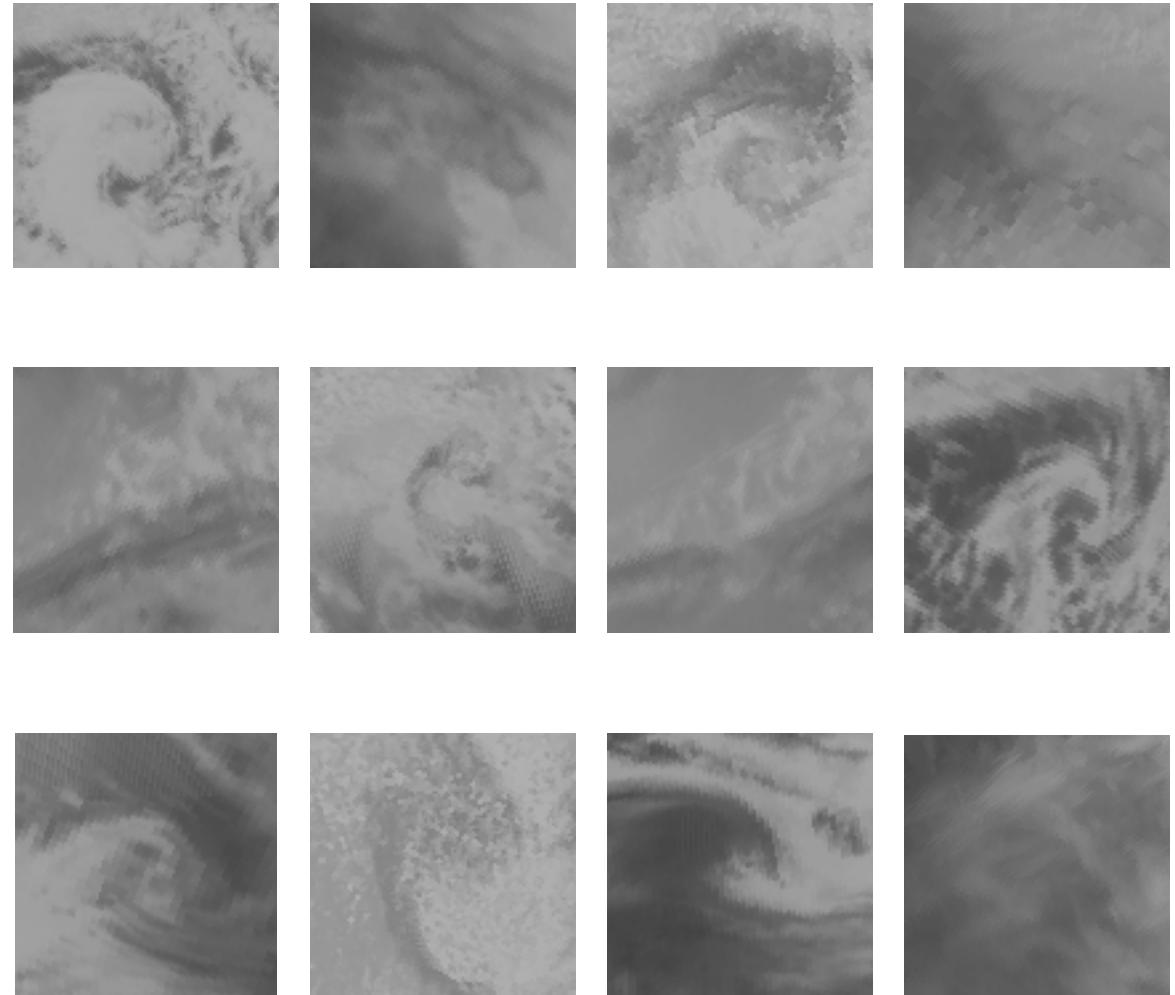
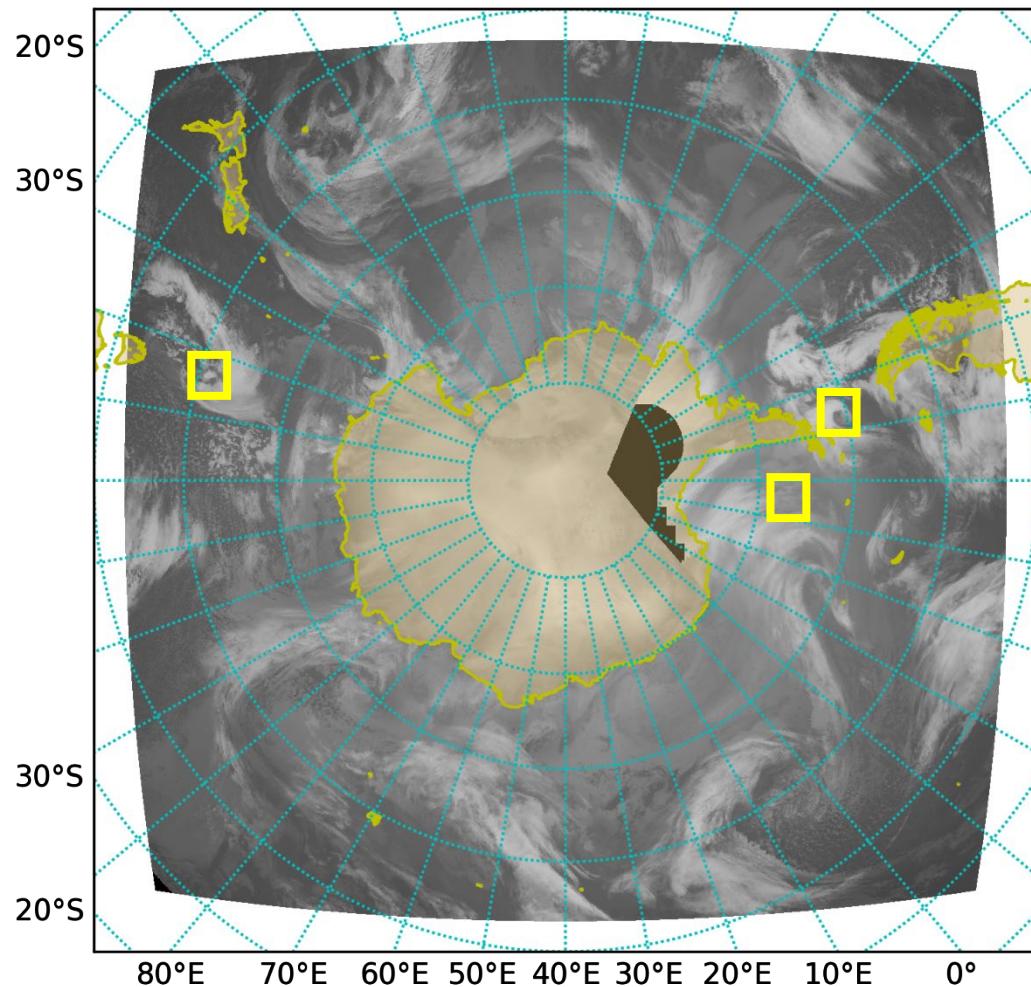
Классификация: исходные данные.



ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Идентификация и построение траекторий ПМЦ

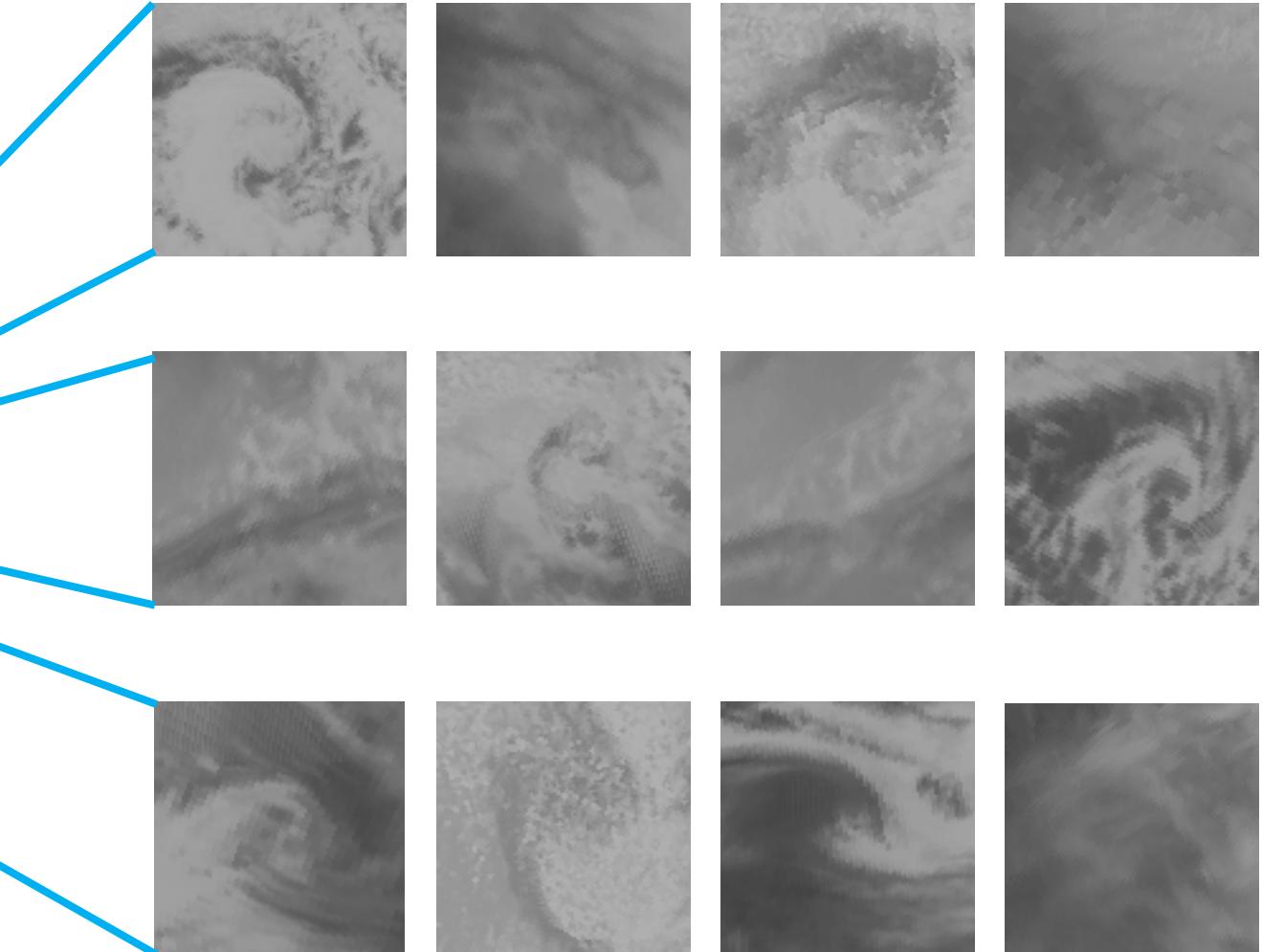
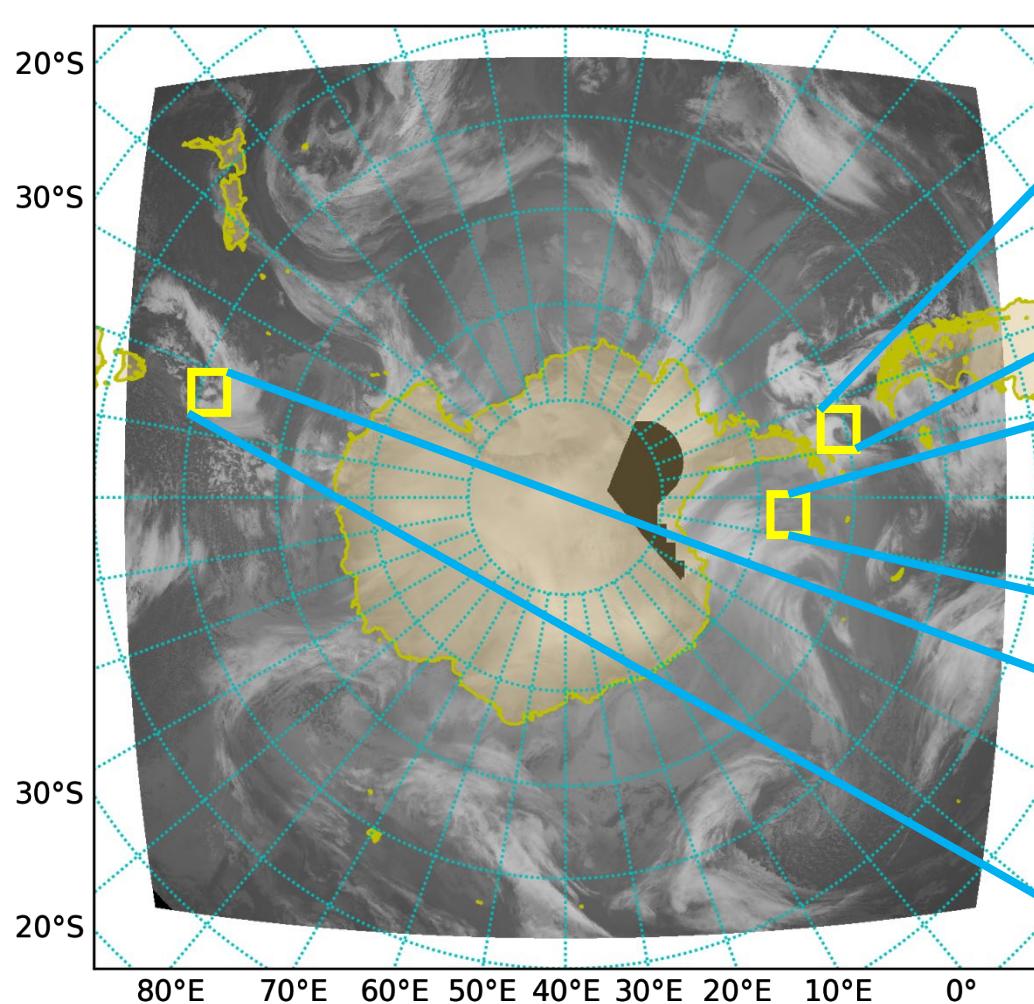
Классификация: исходные данные.



ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

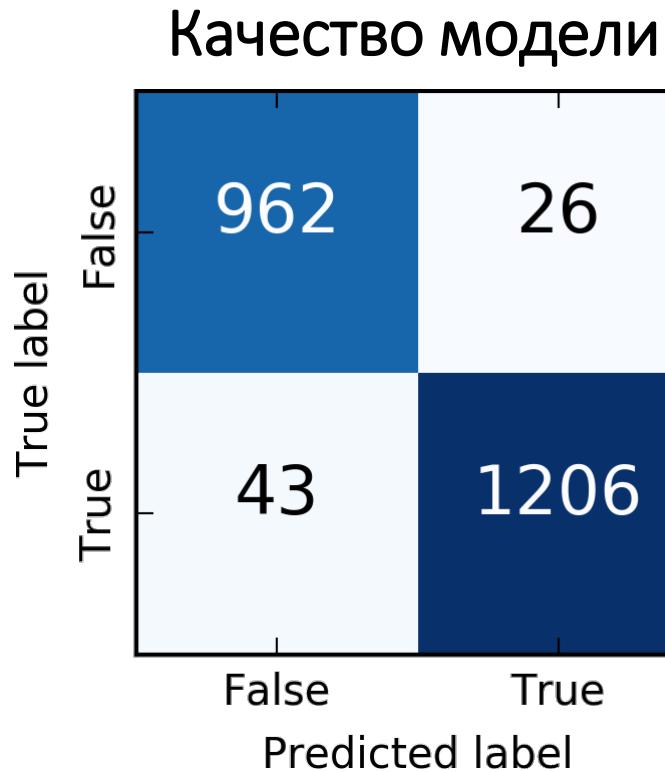
Идентификация и построение траекторий ПМЦ

Классификация: исходные данные.

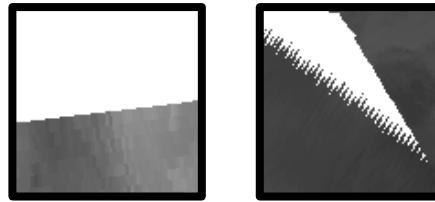


ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

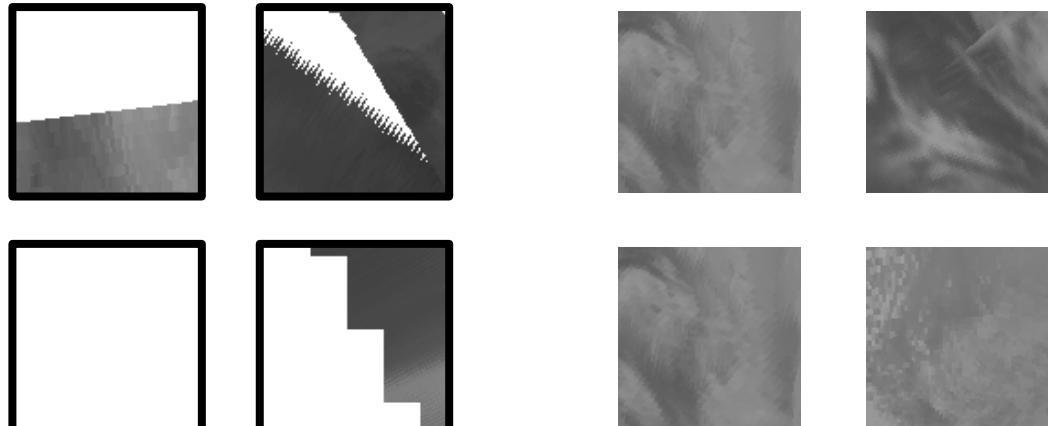
Идентификация и построение траекторий ПМЦ



False negatives – 2%



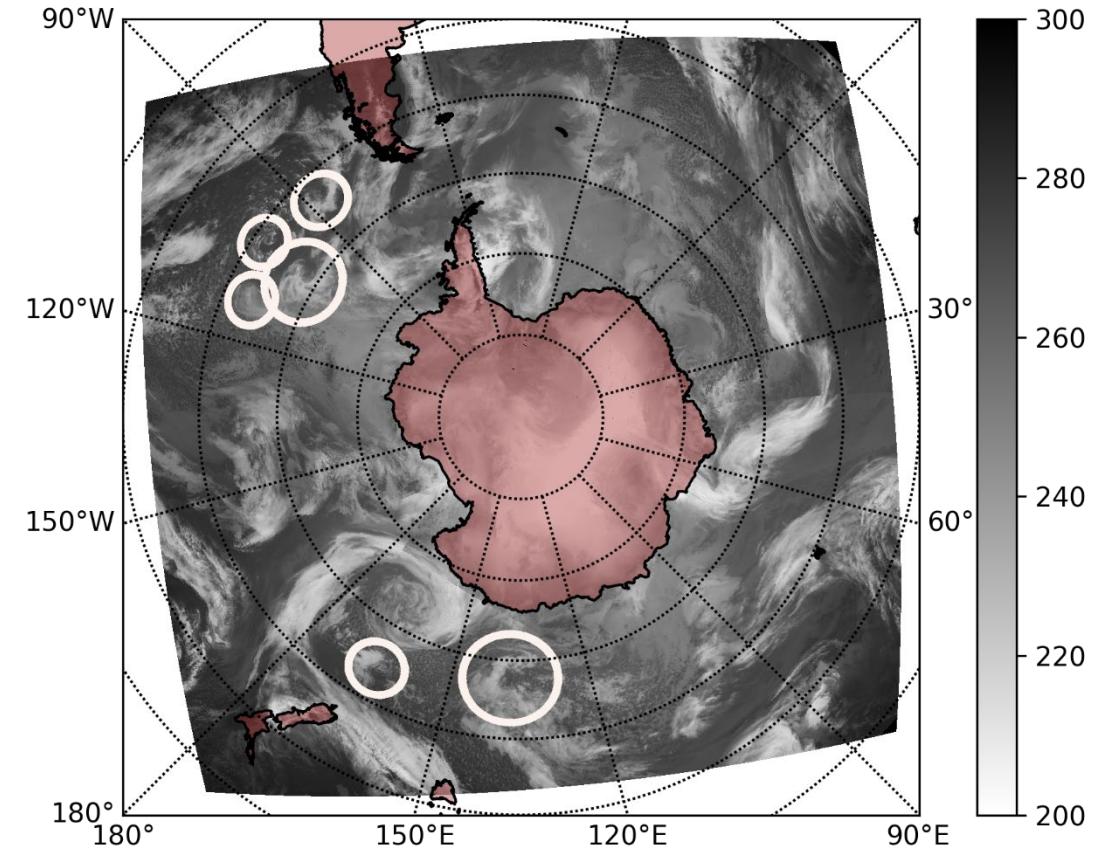
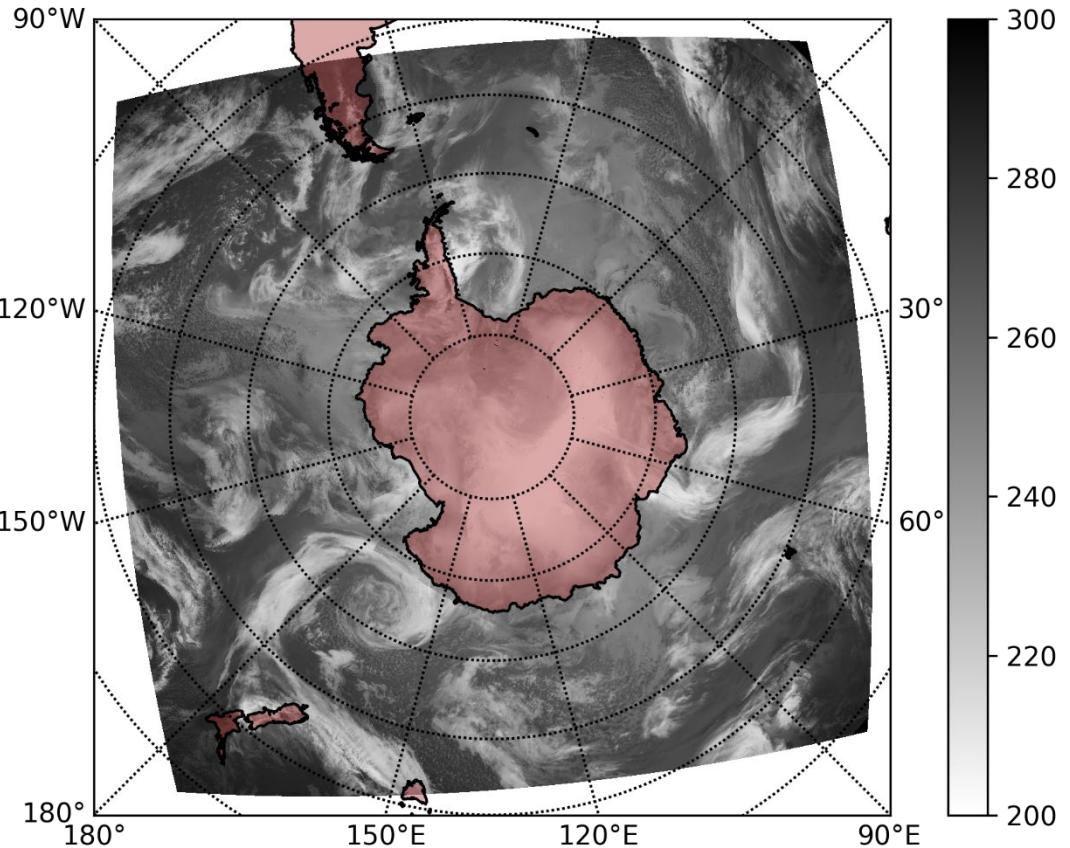
False positives – 1%



ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Идентификация и построение траекторий ПМЦ

Задача №2 - детектирование

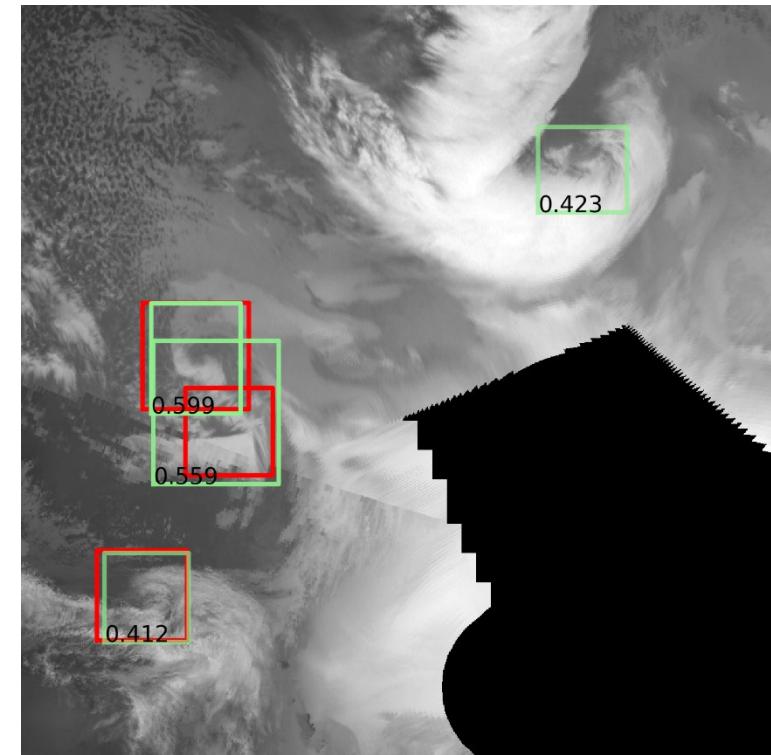
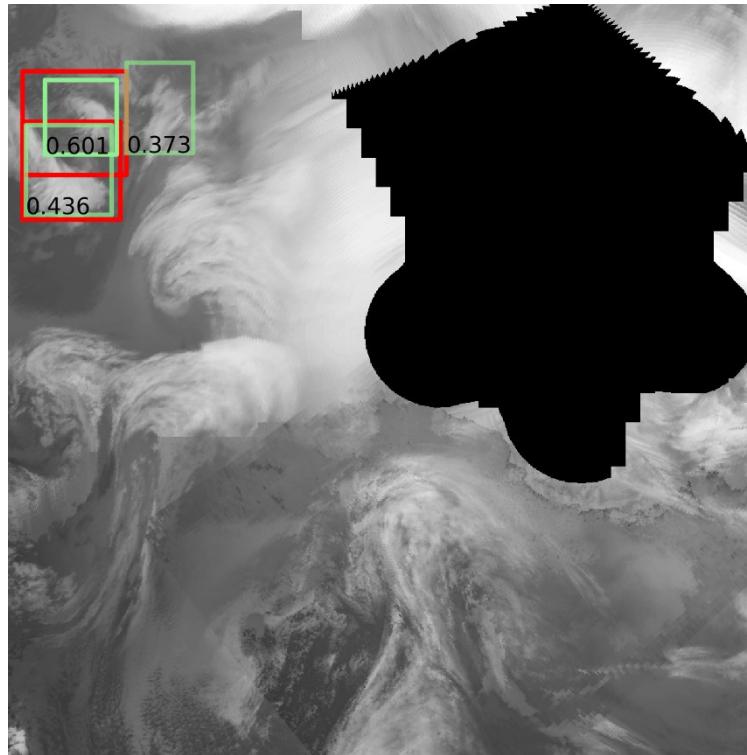
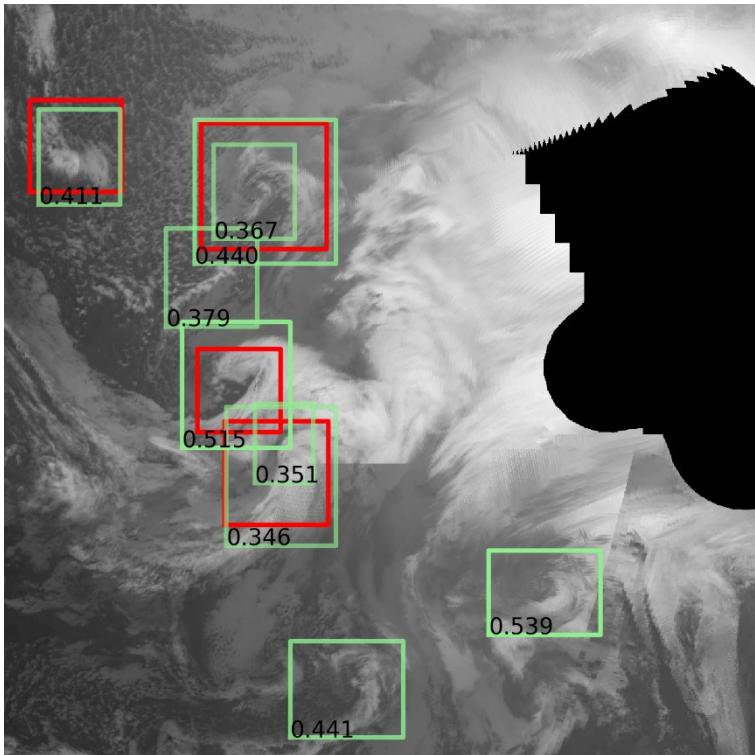


ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Идентификация и построение траекторий ПМЦ

Задача №2 – сегментация

что мы умеем сейчас



ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

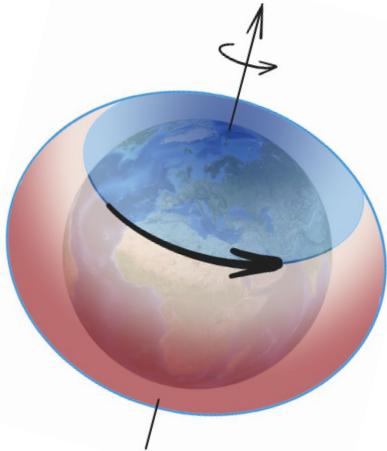
Кластеризация состояний стратосферного полярного вихря (СПВ)

Юлия Зюляева, Михаил Криницкий

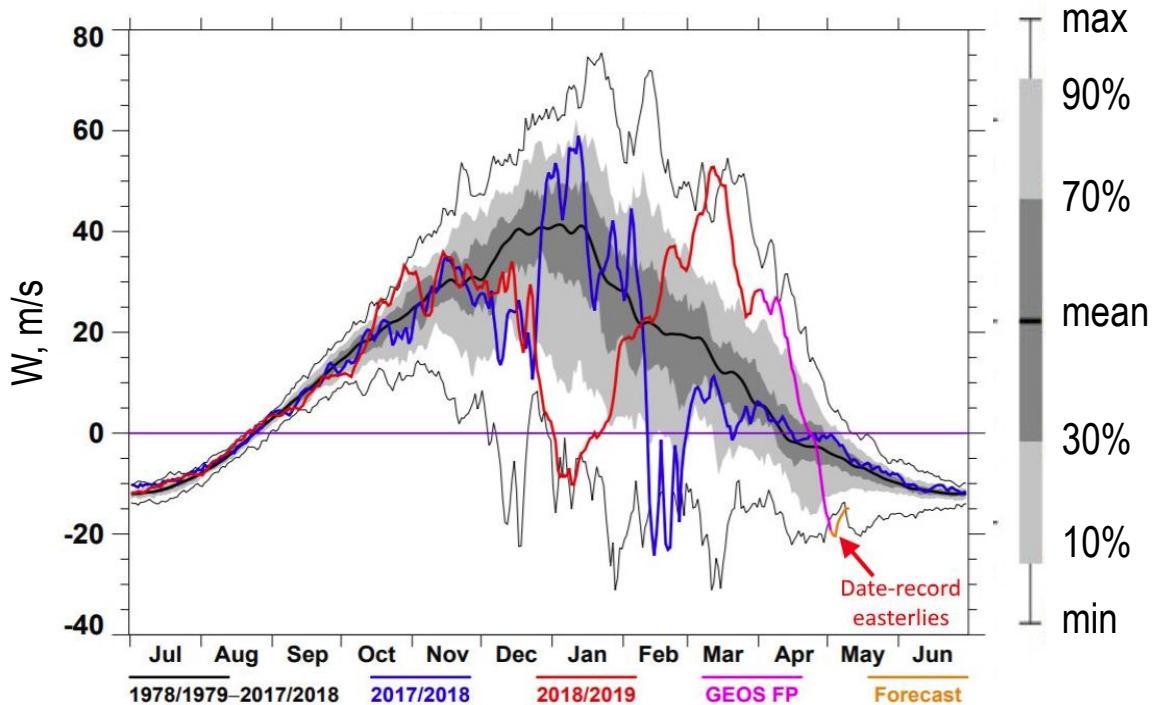
Институт океанологии им. П.П. Ширшова РАН, ЛВОАМКИ

ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Кластеризация состояний стратосферного полярного вихря (СПВ)



Zonal mean zonal wind at 60°N, m/s,
at 10 hPa level (~ 30 km) (MERRA 2 reanalysis data)

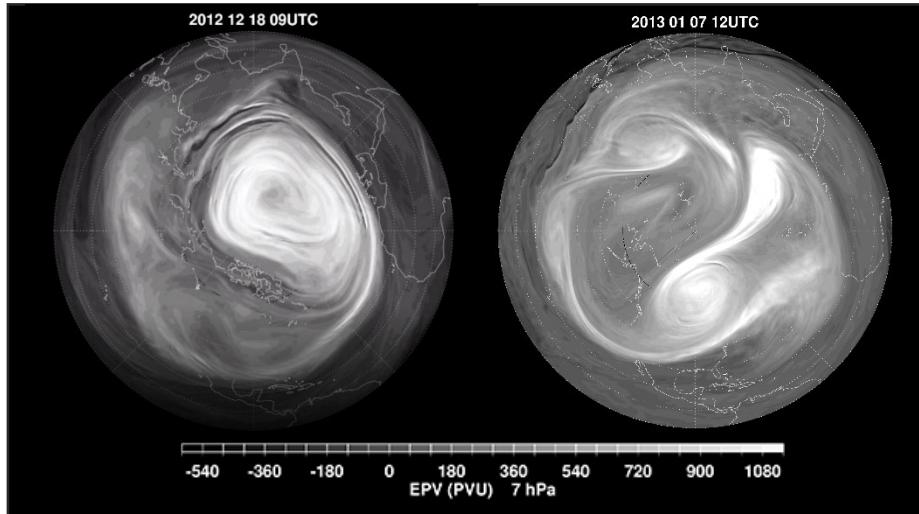


Polar Night Jet: wind speed up to 80 m/s (288 km/h)

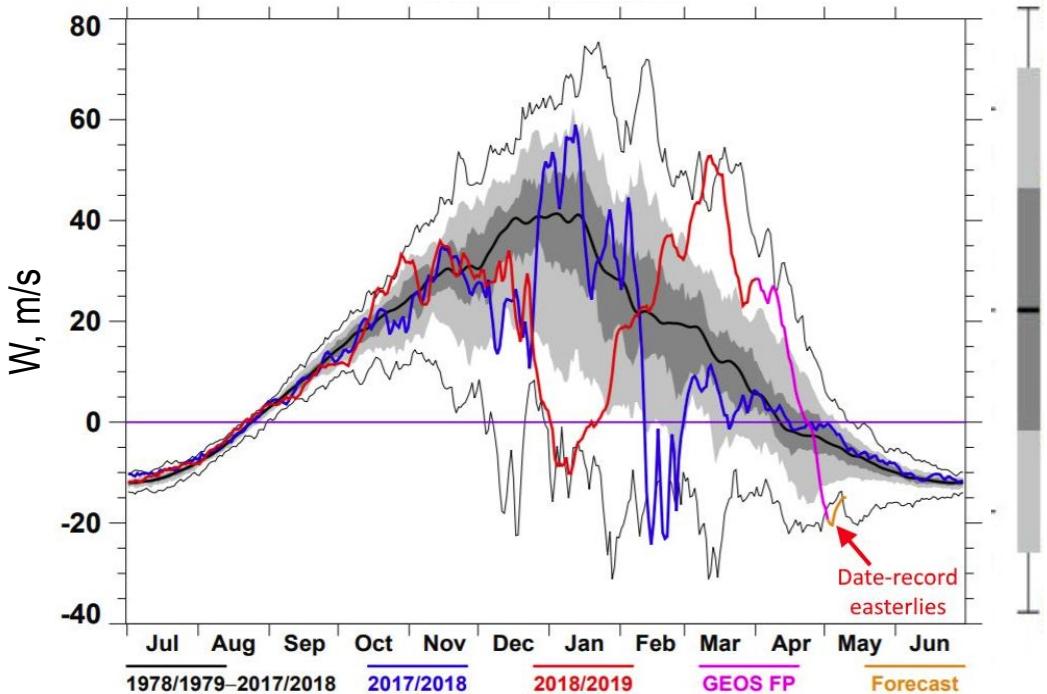
ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Кластеризация состояний стратосферного полярного вихря (СПВ)

Potential vorticity [PVU] at 7hPa level
18-12-2012 (left) and 07-01-2013 (right)



Zonal mean zonal wind at 60°N , m/s,
at 10 hPa level (~ 30 km) (MERRA 2 reanalysis data)



ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Кластеризация состояний стратосферного полярного вихря (СПВ)

What about strong state of PV?

Is it unique? (NO)

Are there different stable shifted strong states? (YES)

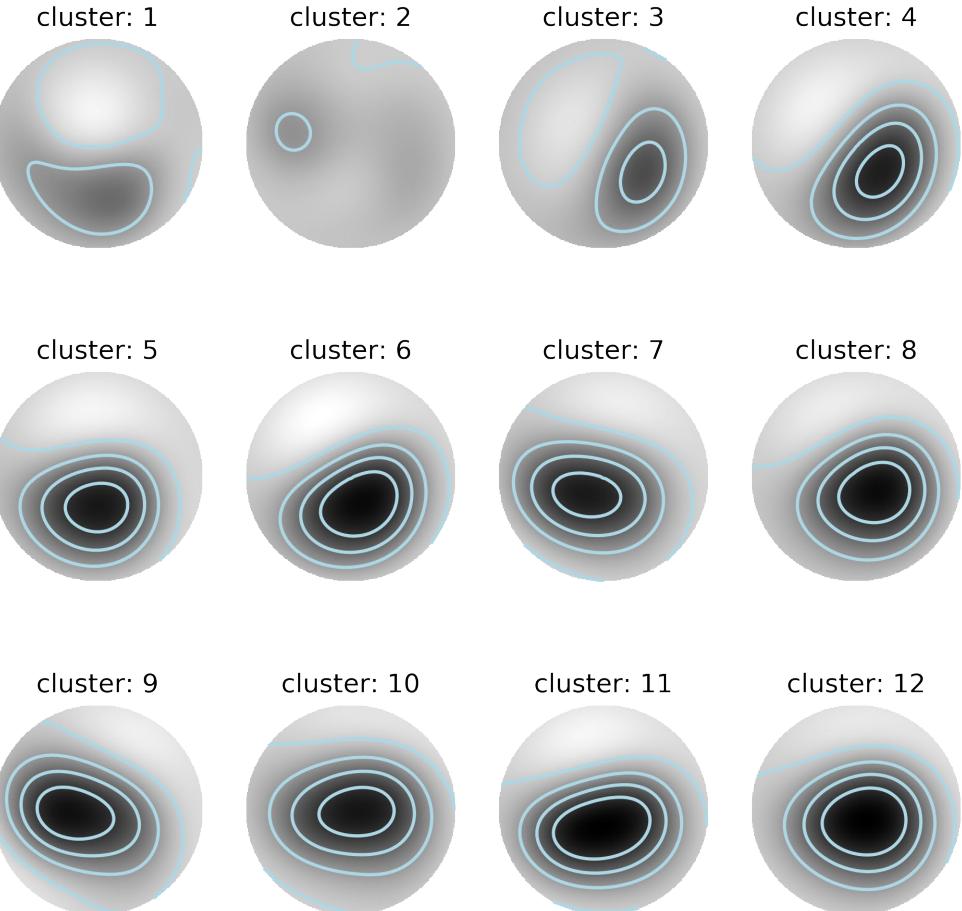
Do they affect troposphere differently? (YES)

ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Кластеризация состояний стратосферного полярного вихря (СПВ)

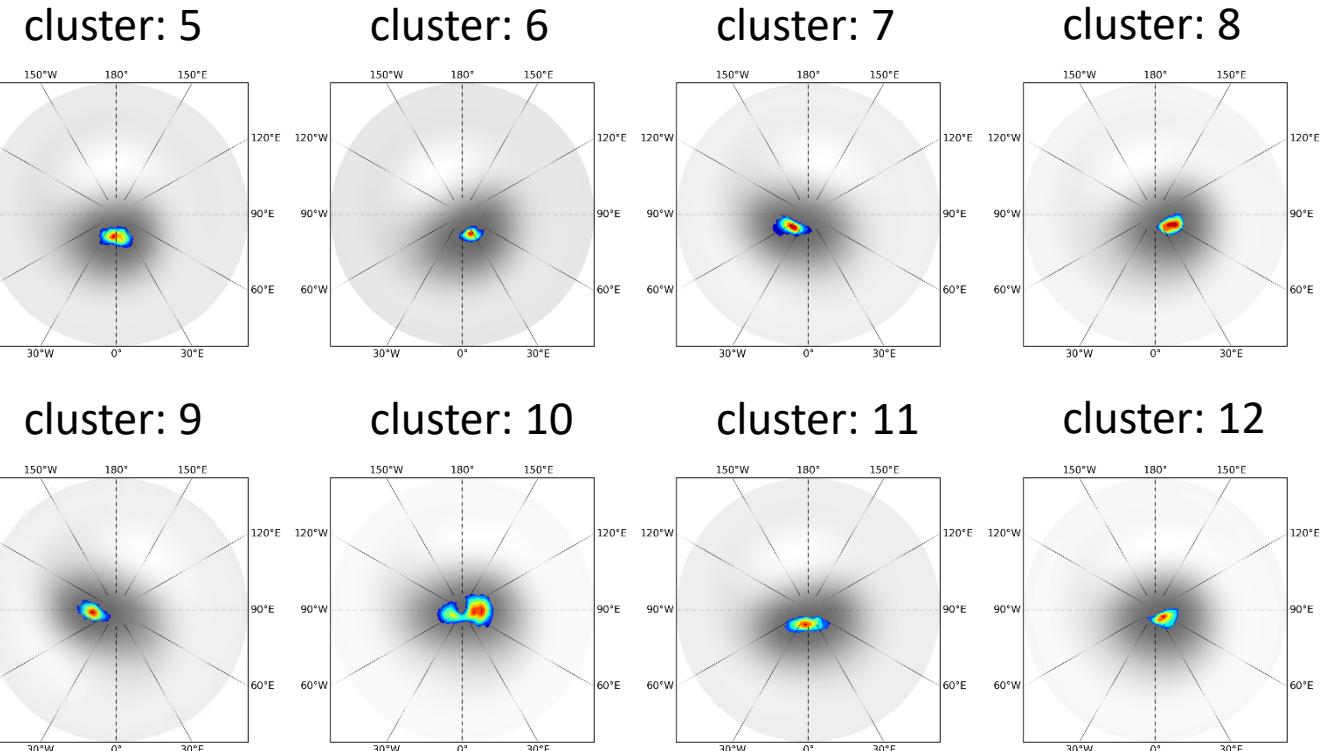
Results and discussion

Composites (HGT field)



Sanity checks:

- Known SSW events of “split” and “displacement” types were clustered
- For strong vortex: variance of center coordinates within a cluster is low



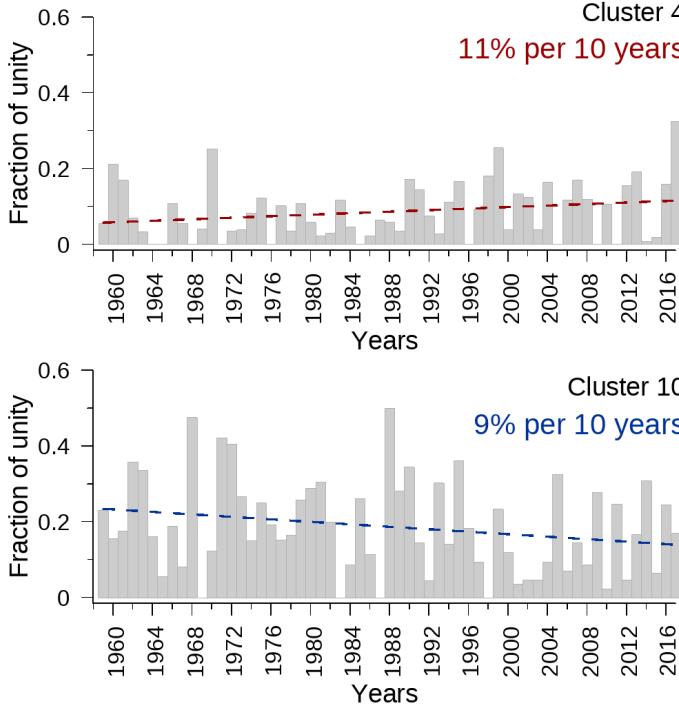
ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Кластеризация состояний стратосферного полярного вихря (СПВ)

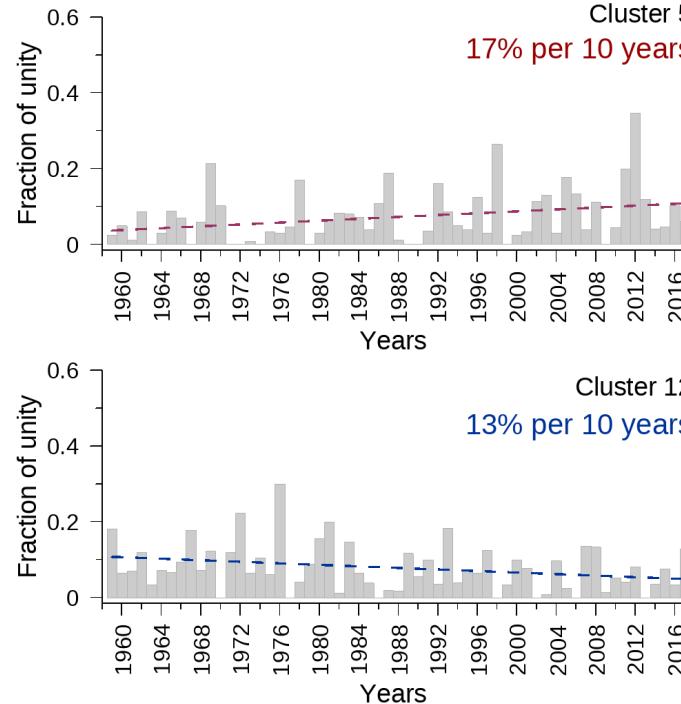
Results and discussion

Linear trends for the frequency of shifted PV states

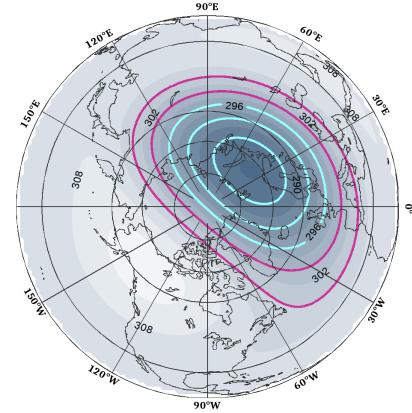
Clusters wintertime frequency (Dec,Jan,Feb), 1958-2017



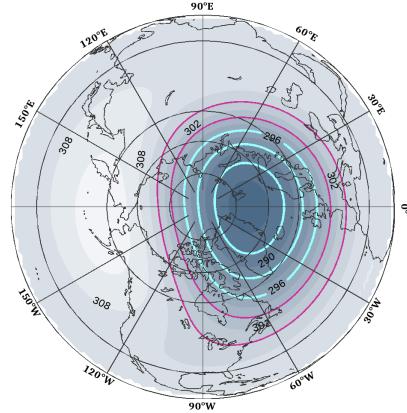
Cluster 5
17% per 10 years



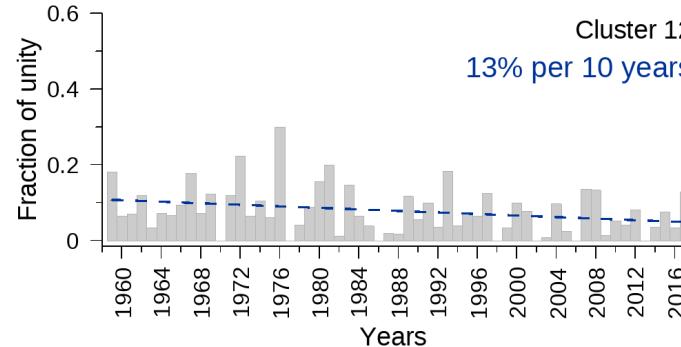
Cluster 4 - 8.6%



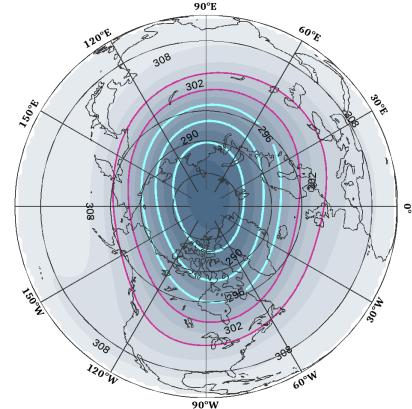
Cluster 5 - 7.2%



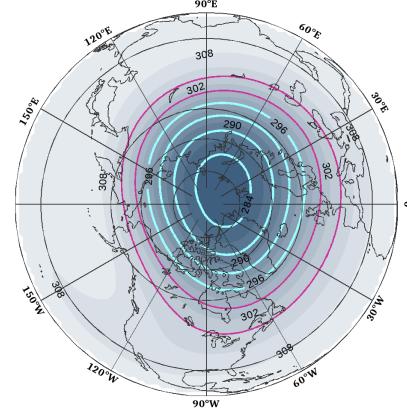
Cluster 12
13% per 10 years



Cluster 10 - 18.7%



Cluster 12 - 7.8%



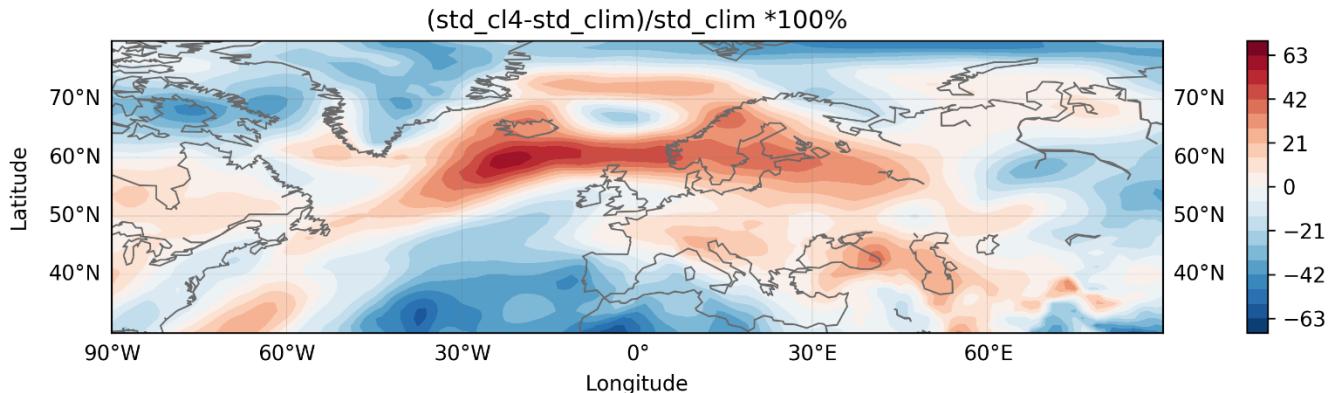
ПРИМЕРЫ ЗАДАЧ НАУК О ЗЕМЛЕ с применением ММО

Кластеризация состояний стратосферного полярного вихря (СПВ)

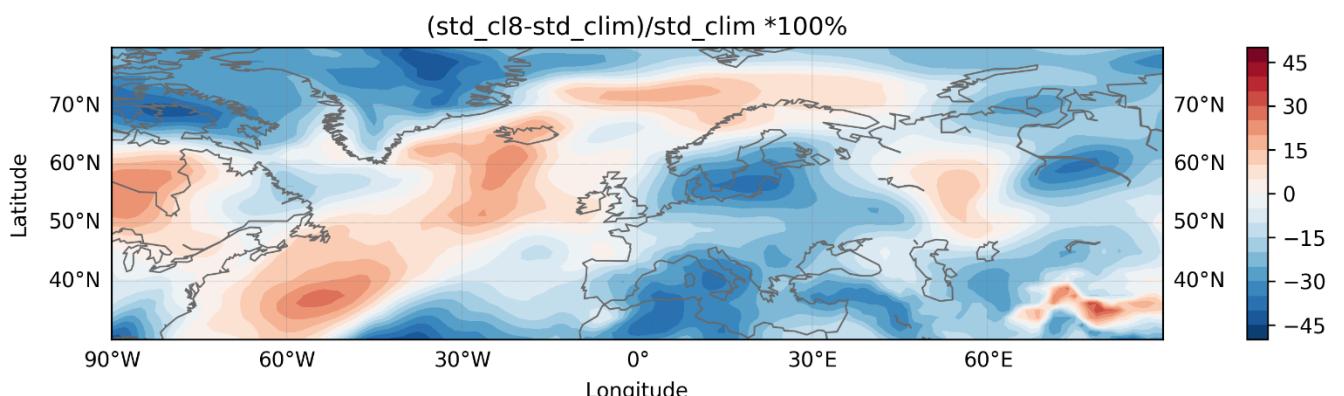
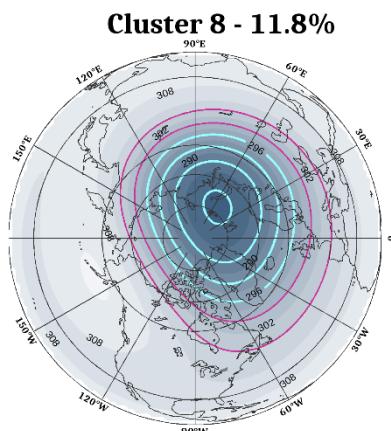
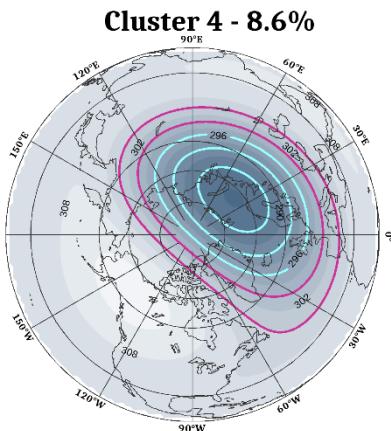
Results and discussion

Troposphere response on the extremely strong PV states

Anomalies of synoptic activity
[% of climatic average values]



(PNJ wind speed >80 m/s;
duration more than 7 days)



ПРИМЕНЕНИЕ СОВРЕМЕННЫХ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ НАУК О ЗЕМЛЕ

- Наукастинг (nowcasting) осадков по радарным данным;
- Классификация видов фораминифер на микроснимке;
- даунскейлинг геофизических полей в численном моделировании;
- уточненная аппроксимация физических величин в задачах полевых измерений: количество облачности, потоки солнечной радиации, приходящие к поверхности;
- задачи фильтрации данных (например, базы данных International Comprehensive Ocean-Atmosphere Data Set, ICOADS).

ПРИМЕНЕНИЕ СОВРЕМЕННЫХ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ЗАДАЧАХ НАУК О ЗЕМЛЕ

- В основном прикладные задачи с понятными результатами;
- непонятные результаты трудно интерпретировать, но они могут быть полезны;
- интерпретация результатов кластеризации обычно еще сложнее, чем интерпретация результатов при «обучении с учителем»;

Все только начинается!